LaJolla
INSTITUTE
P.O. BOX 1434 • LA JOLLA • CALIFORNIA 92038 • ~~PHONE~~

FINAL TECHNICAL REPORT

IDENTIFICATION OF NEW POTENTIAL SCIENTIFIC AND TECHNOLOGY AREAS FOR DOD APPLICATIONS SUMMARY OF ACTIVITIES

APRIL 1, 1985 - JULY 31, 1986

BY
P. HAMMERLING

THIS RESEARCH WAS SPONSORED BY THE
DEFENSE ADVANCED RESEARCH PROJECTS AGENCY
UNDER ARPA ORDER NO.: 3710
CONTRACT NO.: MDA903-85-C-018

The views and conclusions contained in this document are those of the author and should not be interpreted as necessarily representing the official policies, either express or implied, of the Defense Advanced Research Projects Agency or the United States Government.

[redacted]

N O T I C E

LJI-R-86-403

FINAL TECHNICAL REPORT

IDENTIFICATION OF NEW POTENTIAL SCIENTIFIC AND TECHNOLOGY AREAS FOR DOD APPLICATION

SUMMARY OF ACTIVITIES

APRIL 1, 1986 - JULY 31, 1986

BY

P. HAMMERLING
LA JOLLA INSTITUTE
P. O. BOX 1434
LA JOLLA, CA 92038
(619) 587-6000

THIS RESEARCH WAS SPONSORED BY THE
DEFENSE ADVANCED RESEARCH PROJECTS AGENCY
UNDER ARPA ORDER NO.: 3710
CONTRACT NO.: MDA903-82-C-0187

TABLE OF CONTENTS

ADDENDA

"DoD Applications of Electromagnetic Launch (EML) Technologies," LJI-R-85-351, W. Weldon, A. T. Atkin, S. Nozette, and B. Tapley

"Strategic Computing Applications Program," LJI-R-85-348, J. Boris and P. Hammerling

"Four Papers on Robotics: 1. White Paper for the Development of the 'Super Robot', 2. Next Generation of Technology for Robotics, 3. Summary Report on Pipelined Computation of Dynamic Modeling Matrices for Serial Robotic Manipulators, 4. An Assessment of the Development and Application Potential for Robots to Support Space Station Operations," LJI-R-85-349 D. Tesar

"Strategy for Complex Organization Modeling, Planning and Experiment (SCOMPLEX)," LJI-R-85-351, B. J. West

I. 1985 PROGRAM ACCOMPLISHMENTS: SUMMARY

The purpose of the La Jolla Institute's (LJI) technology innovation search program for DARPA is to identify, for consideration of support at universities, research areas either outside of current DARPA programs which could provide the basis for new programs or those, which though they fall within existing programs, offer a significant new approach or idea clearly affecting the program's goals. This search process has been mainly, though not uniquely, confined to universities and makes use of LJI's staff and network of Associates to write reports, convene workshops, etc. to bring both new and overlooked areas to DARPA's attention. ~~Below~~ are summarized some of this year's program activities. Complete copies of the longer reports are reproduced in the Appendix.

II. ABSTRACTS OF REPORTS AND WORKSHOPS

A. "Report on the Polymer Flow Workshop" LJI-R-85-335 [Y. Rabin, Ed. (LJI)]

This report summarizes the workshop held in La Jolla, July 10-12, 1985. Below are listed the meeting's agenda plus a summary.

Polymers in Well-Characterized Flows, (M. M. Reichman, Chairman.)

"Studies of Flow-Induced Conformation Changes in Polymer Solution," L. G. Leal (Caltech, Pasadena).

"Extending Chains by Extensional Flow in Solution: An Approach to Characterize Long-Chain Molecules and Their Interactions," J. A. Odell (Bristol University).

"Theoretical Studies of the Coil Stretching Transition of Polymers in Elongational Flows," Y. Rabin (La Jolla Institute).

"Flow Modification by Polymers," Neil Berman, Chairman

"Dynamics of the Inhibition of Stretching Flows: A Theoretical Study of a Dilute Polymer Solution Beyond the Coil-Stretch Transition in a Stretching Flow of Limited Force," E. J. Hinch (Cambridge University).

"The Orientation and Elongation of Macromolecules in Flow and Related Effects," M. P. Tulin (UC, Santa Barbara).

"Turbulent Flow Interactions and Drag Reduction," J. W. Hoyt (San Diego State University).

"Modern Theoretical Treatments of the Polymer-Flow Interaction," P. Pincus, Chairman.

"Polymer Dynamics in Flow," M. Muthukumar (University of Massachusetts).

"Excluded Volume and Concentration Dependence of Polymer Flow Properties in Dilute Solutions," K. F. Freed (University of Chicago).

"Time-Dependent Calculations of Polymer Dynamics in Flow," Y. Oono (University of Illinois).

"Dynamics of Semi-Dilute Polymer Rods," M. Fixman (Colorado State University).

"Local Conformational Transitions in Polymer Flows," J. Bendler (General Electric).

<u>Polymers in Flow Near Surfaces, Mechanochemistry</u>, (I. Skurnick, Chairman).

"Absorbed Polymer Layers Subjected to Flow," G. G. Fuller (Stanford University).

"Motion of Polymers Near Surfaces," E. A. DiMarzio (National Bureau of Standards).

"The Interaction Between Flow and Chemical Reactivity in Determining Polymeric System Rheology," A. Silberberg (Weizmann University).

Recent experiments on the behavior of polymers in well-characterized flows were reviewed by Leal (Caltech) and Odell (Bristol). These experiments use devices (4-roll mills, cross-slit devices, etc.) in which elongational strain rates are achieved, that are sufficiently high to cause a transition from an unperturbed coil state to a stretched state of the polymers. It has been emphasized that significant flow modification by minute quantities of high molecular weight polymeric additives (<100 ppm)

can be caused only by the stretched polymers since a solution consisting of the latter is no longer dilute (there are many other polymers in the volume spanned by a stretched macromolecule). Such flow modification has been observed by local dynamic light scattering, Leal; and laser doppler velocimetry, Odell; and it is generally believed that the above mechanism is also responsible for other dilute polymer solution phenomena such as turbulent drag reduction, "bathtub vortex" inhibition, anomalous pressures in Pitot tubes, inhibiting the cavitation of bubbles, etc.

The theory of the stretching of polymers by elongational flows has been discussed by Rabin (LJI). It consitutes a first attempt to combine a molecular level description (including polymer molecular weight and flexibility) of the polymers, with a proper description of their coupling to the flow field. The results are in excellent agreement with experiments on the molecular weight dependence of the critical strain rates needed to stretch the polymers in elongational flow.

The consequences of the stretching of polymers for flow modification have been investigated (theoretically) by Hinch (Cambridge) and Tulin (UCSB). Hinch has analyzed the somewhat hypothetical problem (related to fiber spinning) of a dilute polymer solution column that is being stretched by a constant external force. The salient conclusion is that as the polymers get stretched by the local strain rate, their contribution to the fluid stress (momentum transfer) increases dramatically, resulting in a negative feedback that tends to reduce the strain rate in the

fluid to the level where it is barely sufficient to maintain the polymers in their stretched state. Tulin has investigated the anomalous pressure drop in Pitot tubes (in dilute polymer solutions). In both cases, drastic modification of the unperturbed flow (due to the presence of minute quantitites of polymers) was predicted to occur already in the laminar regime.

The phenomenology of flow modification in the turbulent regime (turbulent drag reduction) has been discussed by Hoyt (SDSU) who had emphasized that considerable drag reduction can be achieved by a variety of additives such as polymers, fibers, soaps and even sand. The distinguishing feature of polymers is the unusually low quantity of additive (>0.02 ppm) needed to obtain measurable drag reduction; this, we believe, is a consequence of the fact that a dilute (by weight) solution of polymers becomes quite concentrated (the polymer volume fraction increases by a factor of about 10^4-10^8!) in the stretched state, thus producing a measurable flow modification.

The current state of theory of polymer solution dynamics has been reviewed by Muthukumar (Amherst), Freed (Chicago) and Oono (Urbana). Muthukumar has reviewed the recently developed methods (cluster expansions, effective medium theory) for dealing with the viscosity and other kinetic coefficients of semi-dilute polymer solutions. Although the results are in a reasonable agreement with experiments on quiescent and weakly sheared solutions, they have not yet been extended to the strong flow regime (e.g., the experiments of Leal and Odell). Freed and Oono had focused on the application of renormalization group methods to polymer solution

dynamics. These methods are universally considered as the best theoretical tools for dealing with static properties of polymers; their application to the calculation of dynamic quantities and, in particular, to realistic flow situations is the subject of ongoing research and it remains to be seen whether they will be able to incorporate such important features as polymer inelasticity (at high elongations) and entanglements between the polymer chains.

Results of numerical simulations of the dynamics of semidilute solutions of rigid rods have been reported by Fixman (Fort Collins). They show that the "cage" concept of the Doi-Edwards theory is probably wrong at intermediate concentrations and indicate that the presently popular "reptation in a tube" type description of polymer motions may be inappropriate for the semidilute solution regime.

Going to the more concentrated polymer systems (amorphous solids), an attempt to connect the local conformational dynamics of polymer chains to be observed visco-elastic spectrum in "solid state flows," was reported by Bendler (GE). The conformational potential energy functions for polymers such as polystyrene, PMMA and polycarbonate were computed using quantum (Hartree-Fock type) methods. In view of the fact that intermolecular interactions and constraints were not taken into account in this model, the agreement with experiments on stress relaxation, yielding and crazing, glass transition temperatures, NMR and dielectric relaxation, etc., is quite surprising.

Recent experimental work on the flow of polymers near surfaces (and fluid flow in the presence of surface-adsorbed polymers) has been reviewed by Fuller (Stanford). The apparent contradiction between hydrodynamic and ellipsometric experiments on flow past adsorbed polymer layers (where thickening of the layer is seen in the first and thinning in the later) can possibly be attributed to the fact that hydrodynamic (pressure drop) type experiments are sensitive to the longest chains present, while ellipsometry measures the average width of the polymer layers. The theory of these phenomena was discussed by DiMarzio (NBS) in the context of related problems such as the separation of polymers of different molecular weights by flow through capillaries and the stretching of isolated, surface-adsorbed polymers by shear flows. While some theoretical understanding of the adsorbed polymer distortion can be claimed at present, the problem of flow modification (under these circumstances) is completely open.

Finally, the coupling between chemical reactivity of polymers and their flow-induced conformational state has been analyzed by Silberberg (Weizmann). It has been suggested that chemical reactions can modify viscoelastic behavior and vice versa, that the stretching of polymers can inhibit or enhance chemical reactivity (by altering the number of effectively available reaction sites). In view of the potential applications it is somewhat surprising that no experiments have been done in this area to date.

The report does not contain all the papers presented in the Agenda, however it does contain some papers in addition to those in the Agenda.

B. <u>"Preliminary Report on the Use of Anti-Matter" LJI-R-85-345 [P. Hammerling (LJI)]</u>

We had a few informal meetings in the Summer of 1985 with proponents of an antimatter production and storage program as well as with some JASON members. Additionally, we have been in contact with one of the Los Alamos group. The players in this arena are Bruno Augenstein of RAND, who first brought this to our attention; R. L. Forward, now at Hughes; Professor Gabrielse of the University of Washington; and a team at the Los Alamos National Laboratory. The concepts start with going to CERN to use the LEAR facility, a low-energy, antiproton machine producing some 10^8 $\bar{P}$/sec, further slowing down these particles, and then storing them in a Penning trap. Drs. Augenstein and Forward would like eventually to combine $\bar{P}$'s and positrons to form [redacted] atoms and ultimately to form a cryogenic ball suspended perhaps by laser levitation techniques. Dr. Forward is advocating the consideration of antimatter in propulsion, there being ~44 kton/gram energy content potentially available.[1] RAND has also discussed a possible SDI role for antimatter and has outlined a research program.[2] Neither the internal (Professors K. A. Brueckner and W. B. Thomson and Dr. P. Hammerling) nor the external (Professors S. Drell, F. J. Dyson, N. Fortson, and W. Happer) panel was particularly convinced by the arguments for these applications as justifying a DoD-sponsored R&D program in antimatter production and storage. Such a program would use available data from ordinary matter experiments and theory to help devise the configuration used for antimatter, and thus need not be

excessively expensive at first. Apart from basic physics experiments already proposed or being done at LEAR: inertial and gravitational mass of the antiproton, its g-value, the spectroscopy of the $P + \bar{P}$ analog of positronium, and possibly the Lamb shift of antihydrogen; no really exciting experiment requiring a bottle of antiprotons was identified by the panels. Particle physicists would be interested in a source of polarized antiprotons which could then be accelerated to high energies. Trapped antiprotons could be the first step in achieving this goal, but interesting though this might be, it is not a DoD objective. Probably the most thought and resources devoted to the use of antimatter come from Los Alamos[3,4] where they have also identified [redacted] applications which presumably justify their R&D efforts. Whatever these applications may be, it will take some years of preparation and investigation to accumulate the data required to make an intelligent evaluation. Assuming for the moment that there is a desire to pursue this further, there is some concern that there is no source within the U.S. to replace LEAR if it is no longer available, unless a FNAL accelerator is reactivated. LAMPF II has not yet been approved. Meanwhile the Canadians (Triumph II) and the Russians (UNK) may have considerable ($\sim 10^{16}$ $\bar{P}$/year) production rates in a few years time.

There is consequently a meeting planned for early October to discuss a better production facility within the U.S. The topics to be discussed cover:

1. a. antihydrogen physics

   b. polarized antiproton sources

c. capture of antiprotons in traps

d. low-energy antinuclear interactions

e. gravitational effects of antimatter

2. possible machine parameters and lattice design

3. electron and stochastic cooling needs

4. deceleration techniques (RFQ, electron cooling)

5. injection and extraction

6. positron sources, positron cooling and $\bar{H}$ formation techniques

7. location of facility

The Los Alamos group is putting together a meeting to address the physics and motivation issues at Fermi Laboratory sometime in the Spring of 1986. The demonstration and availability of a transportable source of antiprotons is a near-term goal of the Los Alamos program.[4] They feel that there may be some interest in using antiprotons as probes of condensed matter such as e.g., superfluid helium.[4] Other possibilities may be identified, hopefully in more detail than at present, in the Spring meeting.

Antimatter production costs have been estimated by Rand.[1] The costs depend on the parameter λ = number of $\bar{P}$ produced and collected/number of P in the incident beam, and whether there is self-power and some energy recovery of the P beam. The costs per mg/year are $\lambda = 10^{-4}$: 1.3×10^9\$ (without), 68×10^6\$ (with self-power); $\lambda = 10^{-3}$: \$$33 \times 10^6$\$ (without), 9×10^6\$ (with). $\lambda = 10^{-4}$ "can be done now," $\lambda = 10^{-3}$ is "a difficult near-term goal." The capital costs have been amortized over 40 years in making these estimates.

Before completely shutting the door to consideration of DoD support, it is suggested that the latter meeting be monitored and a classified briefing from Los Alamos be arranged.

References

1. B. W. Augenstein, "Concepts, Problems, and Opportunities for Use of Annihilation Radiation Energy--Near Term RDT&E to Assess Feasibility: An Annotated Briefing," RAND Report N-2302 AF/RC (May 1985).

2. R. L. Forward, "Antiproton Annihilation Propulsion," AFPRL-TR-85-000 (June 1985).

3. L. Campbell, et al., "Basic Research in Atomic Nuclear and Particle Physics," LA-UR-84-3572.

4. M. V. Hynes, "Physics with Low-Temperature Antiprotons," LA-UR-85-1060 (January 1985).

C. "Comments on the Strategic Computing Program at IPTO"
LJI-R-85-347 [S. Amarel (Rutgers)]

These are preliminary comments on parts of the SC program at IPTO. They are based on limited information about the current state of the program, and are intended to provide inputs for planning. Areas discussed include vision, natural language processing, architectures and expert systems. Work on micro/opto electronics, speech processing and infrastructure is not discussed.

A general comment: A key characteristic that distinguishes the core or basic part of IPTO programs from the SC part is that the latter concentrates on information processing systems for performing certain kinds of real life tasks. It is expected that these tasks can be approached by using results of research in AI, in computer systems, and in micro-electronics. Accomplishing these tasks will also require important technological and methodological advances in order to (a) adapt and scale-up known technology, (b) integrate it into complete systems, and (c) solve new problems that are identified in the course of attempting to build these systems. Also it is expected that R&D efforts on these systems will produce working prototypes within a relatively short period of time (2-5 years).

A fundamental issue in R&D on systems is how to define a "reasonable" system decomposition. It is essential that subsystems with relatively strong interactions among them be studied and developed together. In particular, systems that have rich interactions with a physical environment should be developed, to the extent possible, in the context of experiments within that

environment. Also, it is essential to have a clear view of a task (or a family of tasks) during system development. Different specific tasks and different environments may require completely different system configurations (as work on problem solving, or on control systems has shown us). These issues are relevant to the way in which research on systems can be managed.

Let me comment next on specific parts of the program:

VISION

Goal: To develop a system for controlling the motion of a land vehicle on basis of visual information about its environment (as well as information from other sensors).

The main technical problem is to perform vision in motion so that the information extracted from the environment, and assimilated in the system, is adequate to control the motion in accordance with desired task goals.

The main parameters of the problem are:

a. nature and complexity of the succession of scenes encountered
b. nature of terrain and obstacles; avoidance maneuvers needed
c. time constraints; speed desired
d. dynamic characteristics of vehicle
e. structure of goals of the navigation task

The technical subproblems include: feature extraction, identification; internal representation of scences appropriate to task; tracking of objects in visual environment; choice of grain/resolution, focus; obstacle handling; path planning, plan

modification during execution, handling hierarchies of goals; real-time requirements on computer implementation, reliability requirements; development of design environments and of experimental methodology; technology transfer methods from experimental system to engineering prototype.

Current approach: CMU is the main contractor responsible for developing an entire system. Several other (contributing) contractors are concentrating on design issues, subsystem development, and exploration of new computer implementations--all relevant to the task. The intention has been that methods and subsystems developed by the contributing contractors should be "integrated" in the CMU project, and should eventually lead to a more advanced version of a system. Furthermore, the systems developed at CMU are to be "transferred" to engineering/ engineering/demonstration prototypes that are being built and tested as part of the EAO effort in SC.

The contributing projects can be clustered as follows:

a. studies of relevant visual environments needed to design representations and analysis/interpretation algorithms (SRI, AI&DS-Stanford)

b. approaches to dynamic image handling (UMass, USC, Columbia)

c. obstacle handling; spatial reasoning (Hughes, Honeywell, GE)

d. front-level vision and MP implementation (Rochester)

e. advanced vision architectures; parallel algorithms (UMass, MIT).

A suggested change in concept--with implications on the approach: Since the main goal is to develop/explore a system, priority should be given to the development of an entire system by a contractor--vision subsystems, path planning and control of motion, computer implementation, design methodology and technical approach to system transfer. CMU is already doing this. To the extent possible, the collaborating contracts should collaborate with CMU on the CMU system. Some of the contractors (especially those with strong independent views/approaches) should be encouraged to work toward the development of entire new systems. The systems are to be distinguished by the choice of technical parameters of the task. Two or three systems, each addressed to a different task may be appropriate. Such an approach places full responsibility for attaining system performance on a single PI. Of course, task goal and technical parameters of new systems are to be agreed/negotiated with DARPA.

Technical issues to be addressed in future work: Feedback by state of motion on components of the visual system; goal-directed control of focus and resolution; more work on spatial reasoning; increased work on hierarchical planning and plan execution; MP implementations; methodology and software tools for system transfer to other environments.

Relation to EAO-supported work on ALV: I understand that the goals here are (1) to develop a system, largely with state-of-the-art approaches to vision and control, that will reach a certain desired performance (defined in terms of terrain features and speed) within a given time schedule, (2) to facilitate spread

of technology know-how to industry, so that transfer of a demonstration prototype to engineering models (and production) can be achieved effectively, and (3) to establish and maintain working liaison with the services with the goal of adapting and using ALV technology in military applications.

Transfer of work by IPTO contractors to the ALV project can take place via transfer of specific designs and methods, and also via transfer of entire software (sub)systems. It should be a specific goal of system development within the IPTO program to develop mechanisms for transfering a design from one environment to another. Special workshops and conventional publications/conferences are other important means of transfer.

NATURAL LANGUAGE PROCESSING

Goal: To develop systems for interaction in natural language with a Database, or with a specialized Expert System--in the context of Battle Management tasks.

Technical Problems: (1) Intelligent query processing--modeling the user, linguistic interpretation issues in specialized communication modes, language generation, adaptation of characteristics of a NL processing approach to a task (question-answering, asking for advice from an Expert System). (2) Acquisition of knowledge presented in NL text--linguistic issues, using knowledge about the domain of discourse for interpretation, development of internal representatives.

Current Approach: BBN is the prime contractor for the development of a NL query processing system. Contributing contractors are ISI, UMass (language generation, explanation) and Penn (model of

user). NYU is the prime contractor for the development of a NL text understanding system. They are collaborating with SDC on grammatical issues and on the use of domain knowledge in the interpretation process. SRI is a collaborating contractor working on issues of commonsense reasoning and pragmatics. An effort is being made to integrate these efforts and to formulate a transfer plan to Battle Management applications.

A suggested change in approach and emphasis: Put emphasis in development of entire systems. Work simultaneously on an NL component and a specific question-answering or reasoning system that users wish to communicate with. Be specific about the set of tasks for which the system with a NL interface is being built--not only the domain of discourse. In the case of work on the query processing system, develop a realistic experimental environment. The collaborating contractors should work closely with the two prime contractors in system development. The only reasonable integration that I see in this area is via support with concepts and methods that can be used in system design.

Future Work: Possible combination into a single system of acquisition/storage of knowledge communcicated by NL text and interactive query processing. Develop ways of using graphic information together with NL. Work on planning and control tasks that are relevant to Battle Management. Parallel algorithms should be explored.

More support is needed in this area.

ARCHITECTURES

Work in this area is mainline Computer Science. Research is still needed on hardware systems, multiprocessing systems, software and algorithms for parallel systems, and identifications of tasks for which MP architectures are especially appropriate.

Much of the work to date has been hardware-technology driven (VLSI and networking techniques).

Much more work is needed on software, algorithms, design principles, and experimentation with various tasks.

Conceptual work (taxonomies, models of computation) is needed--before embarking on detailed programs of evaluation/benchmarking.

Efforts should be made to develop sufficient software in each case in order to experiment with computations in the following areas:

a. vision: first, low level; then scene interpretation
b. speech: signal processing and then higher level processes
c. implementation of rule systems used in AI
d. NL processing
e. reasoning in AI with heavy use of knowledge bases.

Simulation facilities are useful--not only to test specific designs, but also to encourage the development of frameworks for representing designs and for manipulating them.

Future work is needed on very high level "compiling techniques" for going from a specification of a computational task into a suggested architecture. This is related to other automatic design tasks in AI. Also techniques are needed for automatic task transfer between different architectures.

BRIEF COMMENTS ON EXPERT SYSTEMS:

The designation Expert Systems is too narrow. This effort should cover a variety of knowledge-intensive problem solving systems.

Current thrusts of the program are on:

a. expert system framework/shells

b. knowledge acquisition techniques

c. techniques of reasoning under uncertainty

d. explanation.

More work is needed on handling specifically time and space, on hierarchical systems that combine quantitative and qualitative reasoning, on systems for planning, and on environments for system improvement and refinement.

While "shells" and Knowledge Representation languages are important, an effort is needed to develop methods for guiding the choice of specific Problem Solving system architectures in response to various classes of tasks/problems.

The system developments in this program should be made in the context of specific problems relevant to the ALV, BM and PA project areas. This is essential not only for testing the systems, but also to identify possible system features needed to handle these problems.

D. Meeting on "Low-Dimensional, Quantum-Well Structures," a Technical Interchange Meeting Held at DARPA 2-3 October 1985

This meeting brought U.K. and U.S. scientists together with DARPA and service representatives to discuss developments in quantum-well structures, particularly how the U.K. program in this area might interface with DARPA's interests.

Professor J. L. Beeby of the University of Leicester summarized the U.K. program sponsored by the Science and Engineering Research Council (SERC), in both one- and two-dimensional structures. The materials considered are mainly the III-V, II-VI. The program started in 1984 and is funded at a level of £6 M/year for 5 years. Dr. Beeby's remarks are summarized below. (An overview of the program is given in the La Jolla Institute Report LJI-R-84-286.) D. Chemla of AT&T Bell Laboratories reported on work in collaboration with D. Miller on the "Physics and Application to Optoelectronics of Room Temperatures Excitoric Resonance in Quantum Well Structures." Dr. M. Pepper of the Cavendish Laboratory discussed his recent work on "Transport in Narrow and Short GaAs FET Structures" (Dr. Pepper is a co-author with K. van Klitzing of the paper reporting the discovery of the quantized Hall effect for which the latter was awarded the 1985 Nobel prize in Physics); his paper is summarized below. Dr. G.A.C. Jones of Cambridge University discussed "Advances in Electron and Ion-Beam Fabrication Techniques for Low-Dimensional Structures" and Dr. D. Anderson of the Royal Signals and Radar Establishment (RSRE) discussed their programs and fabrication methods. Dr. L. Esaki (Nobel Laureate in Physics) discussed the IBM research in two-dimensional

structures while Professor D. Tsui of Princeton University gave a theoretical presentation on "Size Effects on Two-Dimensional Transport." (Dr. Tsui is known for his work, following Klitzing, on the fractional quantized Hall effect which is intimately related to the reduced dimensionality). Dr. Beaumont of Glasgow University outlined his group's fabrication techniques in achieving "nano-lithography" on the scale of 10 nm (100 Å or 10^{-2} µm). He speculated that this could bring him in the range of electronically-simulating neural networks. Professor S. D. Smith's work at Heriot-Watt University on optical computing was summarized by Professor Stradling; this work is considered to be the most advanced in Europe in this subject. Professor Smith's comments are given below; due to his viewgraphs not being available, we also give a copy of a review article which covers many of the same topics. A very interesting presentation was given by Professor R. L. Byer of Stanford University outlining his ideas relating to the study and production of synthetic nonlinear materials; his slides are reproduced below.

The U.K. group is preparing "white-papers" in this area for consideration. However, a more intensive follow-through effort will be needed before a program start could be initiated.

UK PROGRAM

Professor J. L. Beeby
(LDS Coordinator, Science & Engineering Research Council)
University of Leicester

Introduction

The Science and Engineering Research Council's (SERC) programme on Low Dimensional Structures (LDS) has three main general aims:

1. to study an exciting new area of science
2. to underpin the U.K. semiconductor industry
3. to provide needed trained manpower.

The need for an initiative in LDS was first formally noted at a meeting of the Semiconductor and Surface Science subcommittee in September 1982. By November 1983 a report had been prepared on behalf of the Physics committee proposing a programme costing about £30M over five years. This was subsequently supported, but no funds allocated, by the Science Board and the Council. The first major grants were approved in June 1984 and a firm funding line agreed in October 1985.

The Programme

The LDS programme was to build on background facilities already established by the Engineering Board of SERC which included, for example:

MBE:	Glasgow, UMIST, City of London
Si fabrication:	Edinburgh, Southampton
e^--beam lithography:	Rutherford Laboratory
MOCVD, LPE:	Sheffield

The proposal suggested that three new growth systems should be established in each of the first three years of the programme and that they should be in, or in close proximity to, the department undertaking the research, which would be mainly physics. It was thought that a single growth system would be able to support an average of four research projects at any one time.

The LDS community thus established was estimated to number about 60 academics, 60 research associates, 30 technicians, and 30 research students by 1988. In practice it is already much larger. The materials systems were expected to include III-V's, GaInAs Al PSb, II-VI's, both wide and narrow gaps, IV's, SiGe alloys and dopants, and also metals and insulators in conjunction with semiconductors. The structures of interest included both two dimensions and one dimension.

Management

The management of the programme will be by SERC with funds allocated by "peer group" committees. General policy advice will come from a Programme Advisory Group consisting of industrialists, government representatives (defence and trade) and scientists and engineers. The overall direction will be in the hands of a coordinator who will also be responsible for international cooperation.

EXPERIMENTAL SEMICONDUCTOR PHYSICS AT CAVENDISH LABORATORY:
TRANSPORT IN NARROW AND SHORT GaAs FET STRUCTURES

Dr. M. Pepper
Cavendish Laboratory

We have fabricated GaAs-AlGaAs modulation doped heterojunction FETS where two types of gate were used. The first was ≈800Å long, whereas the second was 15 μm length with a ≈1 μm gap in the centre.

The first type of device showed very interesting current-voltage characteristics when the channel under the gate was depleted. The current, I, varied with voltage, V, as V^2 indicating space charge limited behaviour. This was expected as rough calculations indicated that the injected charge density was greater than the intrinsic charge density.

The differential of the current was investigated as a function of source-drain voltage and structure was found when this voltage corresponded to the optical phonon energy. This showed that some electrons were travelling across the channel picking up the applied voltage and then emitting an optical phonon when the acquired energy corresponded to the threshold for this process.

In the second device the action of the gate was used to squeeze the conducting channel. The conductivity of magneto-conductivity were measured as a function of temperature below 4.2°K. It was found that both quantum interference and electron interaction effects were present. Analysis of the negative magneto-resistance showed that the mechanism of phase relaxation was by the scattering of electrons from electromagnetic fluctuations arising from fluctuations in charge distribution. The

temperature dependence showed that electron interaction effects were present and all corrections were one dimensional. This was not surprising as the analysis showed that the width of the channel was 500Å, sufficiently small that one-dimensional behaviour is expected.

Results were presented on the transition to hopping behaviour as the carrier concentration was reduced.

HERIOT-WATT PROGRAMME IN NONLINEAR OPTICS AND OPTICAL COMPUTING

S. Desmond Smith
Department of Physics
Heriot-Watt University
(presented by Professor A. Stradling)

Professor Smith sends his apologies for being unable to attend this meeting and has asked me to read this statement of the interest at Heriot-Watt University in current advances in multiple-quantum-well material fabrication and properties.

As you may know, the major semiconductor research programme at Heriot-Watt concerns the understanding and development of optically-bistable and "transphasor" devices, whose range of applications includes the currently very exciting area of optical computing.

Following research and development during the 1960s on the so-called Spin-Flip Raman laser--for a tunable infrared laser source--concern with the transmission of CO laser radiation at 5.5 um wavelength, through indium antimonide, lead to the reduction of the giant nonlinear refraction of this material at frequencies just below the fundamental bandgap.

Dr. Chemla's present co-worker--David Miller, then working at Heriot-Watt--measured nonlinear refractive coefficients of up to 10^{-3} cm^2/Watt, this being some eight or nine orders of magnitude greater than the Kerr-medium nonlinearity of CS_2 which is used effectively as a standard.

1. The resonance enhancement of the InSb coefficient near the band edge is shown in this overhead. Note

particularly that the high n_2 is achievable in a spectral region where the semiconductor linear absorption coefficient is less than 100 per centimetre of material. This combination of properties made this small-gap semiconductor particularly suitable for Fabry-Perot optical bistability and indeed the first demonstration of this phenomenon with continuous wave lasers as made in Edinburgh during 1979.

2. Here we show the major range of transmission and reflection characteristics achievable with nonlinear Fabry-Perot etalons. The possibility of either hysteretic (optically-bistable) or nonhysteretic (transphasor) responses is allowed by the initial detuning, of a Fabry-Perot resonance of the semiconductor cavity, with respect to the operational laser wavelength. [See Figure 1 (Reprint below)].

3. The overhead shows such detuning being varied by adjusting the orientation of the sample with request to the laser beam. Note the power level required for switching--of the order of 10 mW for a laser spot size around 50 μm.

   The variety of response characteristics means that one has here in principle both the optical switches and the optical logic gates that would be required as components of an optical digital computing device as well as optical

amplification and power stabilisation analogue performance regions.

4. We have been successful during the past five years in demonstrating the range of logic gates, of switching the transmission of the cw CO holding beam with a number of sources including 35 psec YAG pulses, modulated He-Ne and diode laser beams, and incoherent camera flashes. [See Figure 3 (Reprint below)].

We have also performed the optical equivalent of Jack Kilby's 1958 experiment at Texas Instruments in that we have operated with two independently switchable pixels on a single InSb slice, and have used the switched output from one to cause the switching of the second.

Whilst InSb has a dramatically high nonlinearity there are disadvantages regarding the device application of the InSb/CO system. Firstly, the necessity of cryogenics--we operate a liquid nitrogen temperatures. Room temperature operation at 5.5 μm is precluded because the InSb bandgap falls to 7 μm at 300°K. The free carrier intrinsic absorption is likely to prohibit low power cw operation. We have in fact observed pulsed, 10.6 um bistability at room temperature in InSb. Secondly, use of infrared radiation makes initial development of optically coupled circuits more awkward than would a visible system. More importantly though, the spot-size diffraction limits at longer wavelength reduce the potential packing-density

for the arrays of switches or gates that are being envisaged.

Thirdly, we have shown the carrier diffusion length in our devices to be some 60 μm, and computational studies indicate that switching cross-talk is significantly affected by the diffusion at up to five times the diffusion length--for linear arrays of pixels.

Given sufficient commitment these disadvantages might well be reducible to an acceptable level. Perhaps more significant at the moment, however, is the difficulty in acquiring large area (say 1 cm square) samples of accurate spectral and spatial uniformity and high surface parallelism.

5. As an alternative strategy we are currently also investigating the use of Fabry-Perot-like dielectric multilayer interference filters. Similar response characteristics are achieved as in the InSb case as shown on this overhead. Now, however, we have cw Argon-ion holding beams in the visible (514 nm green), room temperature operation, and the back-up of a developed technology for fabricating large <u>uniform</u> plates. Once again note that cw holding beam levels of around 10 mW are required for 50 μm spots. There is, however, one significant difference between the two devices. InSb operated, using a nonlinearity of electronic origin and with a characteristic time scale of 100 nsec: the

filters operate using the thermal nonlinear refraction of the central (ZnSe) spacer region, with thermal timescales typically around tens of microseconds.

In either case, for switching arrays one almost certainly requires holding power levels at least two orders of magnitude smaller than quoted above. To achieve this one needs firstly material optimisation for higher nonlinear coefficients, secondly optimisation of the sample structure, and thirdly reduction of the radiation spot-size.

Thus, for example, switching powers down to around 2 mW have been achieved at spot-sizes of 20 μm and both theory and experiment predict continued reduction down to diffraction limited regimes. Sample pixellation--with free-carrier or thermal isolation--is also required in this limit, as is a uniformity pixel-to-pixel, on this spatial scale, over the complete sample.

Sample structure optimisation is unlikely to produce dramatic improvements in bistability power levels. However, thin films of materials such as InSb, or containing InSb, surrounded by suitable high-reflectivity coatings could improve the situation for electronic-nonlinearity based devices. And, surprisingly, relatively thick (perhaps 10 μm) accurately flat and reproducible layers in thermal devices are predicted to give lower power nonlinear responses that the submicron spacers used at present. Vacuum thermal deposition techniques restrict

the acquisition of thicker spacers of good uniformity. Turning to the nonlinearity itself--in particular to electronic nonlinearity--this is the area in which MQW materials have been mooted as being of most significance. It is not yet clear that the isolation of the excitons from the GaAs bandgap, by using MQM structures, does indeed give the anticipated improvement in room temperature all-optical bistability conditions. The point has recently been discussed by Ovadia and co-workers from Hyatt Gibbs' group. In principle, however, the presence of a small absorption in the tail of a spectral region of high joint-density of optically-coupled states and high oscillator strength is required for strong nonlinear refraction.

6. The basic physical process for electronic nonlinearity is shown in this overhead. Given a two-level system, this shows a homogeneously-broadened absorption and the saturation of this absorption at high pumping conditions. By causality the two-level dispersion, shown here, also saturates so that at frequencies below the transition frequency there is a reduction of the refractive index. For bulk semiconductor systems, carrier generation at a frequency in the tail of the band again causes a partial saturation of the interband absorption--as could be probed at low irradiance over a range of frequencies--and again $\bar{n}$ is reduced in the band gap region. The

simultaneous creation of valence holes complicates the situation but does not qualitatively alter it.

7. Here we show the scaling of the nonlinear index at high temperatures (valid at 77°K in InSb). Note the $1/E_g^3$ factor. Narrow gap materials are strongly favoured for high n_2 and as a consequence for low irradiance bistability. This parameter I_c characterises the switching irradiance. These formulae apply in the case that the carrier recombination is trap dominated rather than dominated by radiative recombination or Auger transitions. T_1 is the recombination lifetime. The frequency factor J contains the band-gap enhancement and of dependence on the cavity structure.

Overhead 1 is shown again here to demonstrate the fit of the theory and experiment using this formula.

8. Here the n_2 is modeled for both bulk material and for the quantum well step (ignoring excitonic effects). For the same linear band tail absorption and same T_1 one would achieve similar coefficients in the two cases. It is appropriate therefore to ask what tailoring of the carrier recombination time is achievable in MQM structures.

Could one either reduce the switching power by increasing T_1 (at the expense of switching rates) or alternatively

improve the switching time and carrier diffusion lengths by reducing T_1?

9. Here a variety of band edges are schemed. In practice the near-gap joint-density of states in small gap semiconductors is actually low, because the effective mass of the conduction band is small. However, the interband oscillator strength is high (proportional to E_g^{-1}) and is dominant in creating the strong nonlinearity. An apparently sharper edge is certainly acquired either in the presence of excitons or of step-like two-dimensional sub-bands or of one-dimensional London sub-bands. However, the near edge feature must contain sufficient total oscillator strength if it is not to become saturated at irradiance levels below that required for bistability, and if advantage of the feature is to be taken.

<u>In Summary</u>

It may prove that the advantage of MQM materials lies not in the edge features but in the fact that the position of the edge can be tailored to match available high power cw (or perhaps pulsed) laser sources. In this context we would currently be particularly interested in materials of high uniformity and 5.5 μm edges at either 77°K or room temperature, and in materials with near-514 nm edges at room temperature.

We are also interested in the fabrication techniques used for MQM materials, as applied to the production of very high quality "thin" and "thick" films of single crystal materials--films of thickness varying from perhaps 100 nm up to 50 nm.

14. Vinogradov, A. V. & Shevagurov, V. M. Sov. J. quant. Electr. 13, 303-306 (1983).
15. Rosen, M. D. et al. Phys. Rev. Lett. 54, 106-109 (1985).
16. Gudzenko, L. I. & Shelepin, L. N. Sov. Phys. Dokl. 10, 147-149 (1965).
17. Suckewer, S., Zimmerman, G., Lokke, W., Chapline, G. & Wood, L. Bull. Am. Phys. Soc. 17, 972 (1972).
18. Pert, G. J. 1st Natn. QE Conf. (Manchester 1973).
19. Pert, G. J. et al. in Laser Techniques in the Extreme UV (eds Harris, S. E. & Lucatorto, T. B.) (AIP, 1984).
20. Suckewer, S., Keane, C., Milchberg, M., Skinner, C. H. & Voorhees, D. in Laser Techniques in the Extreme UV (eds Harris, S. E. & Lucatorto, T. B.) (AIP, 1984).
21. Seely, J. P. et al. Centre for Space Research, Preprint (1985).
22. Jaegle, P. et al. J. Phys. B (in the press).
23. Jaegle, P. et al. in Laser Techniques in the Extreme UV (AIP, 1984).
24. Vinogradov, A. V., Sobelman, I. I. & Yukov, E. A. Sov. J. quant. Electr. 5, 59-63 (1975).
25. Dixon, R. H. & Elton, R. C. J. Opt. Soc. Am. 1, 232-238 (1984).
26. Apruzese, J. P., Davis, J. & Whitney, K. G. JAP 53, 4020-4027 (1982).
27. IEEE J. Quant. Electr. Spec. Issue on Free Electron Lasers QE 19 (1983).
28. Goldstein, J. C., Newnam, B. E., Cooper, R. K. & Comly, J. C. in Laser Techniques in the Extreme UV (eds Harris, S. E. & Lucatorto, T. B.) (AIP, 1984).
29. Dobiasch, P., Meystre, P. & Scully, M. O. IEEE J. quant. Electr. QE 19 1812-1820 (1983).
30. Dixon, R. H., Seely, J. F. & Elton, R. C. Phys. Rev. Lett. 40, 122-125 (1978).
31. Harris, S. E. Opt. Lett. 5, 1-3 (1980).
32. Wong, J. C., Caro, R. G. & Harris, S. E. Phys. Rev. Lett. 51, 767-770 (1983).
33. Robinson, C. A. Aviation Week and Space Technology 23 February, 23-27 (1981).

Lasers, nonlinear optics and optical computers

S. D. Smith

Department of Physics, Heriot-Watt University, Edinburgh EH14 4AS, UK

Devices whose optical properties change with light intensity have opened the path towards the optical computer. Optical switches and signals can now undertake all binary logic operations and complete optical digital circuits have been constructed.

FROM the earliest days of the laser, the large intensities available in laser beams permitted irradiances of the order of MW cm^{-2} incident on material by simple focusing. With such fluxes the amplitude of the oscillating electric field, E, readily becomes comparable to interatomic fields such that the optical properties of the medium vary with the passage of the wave. The polarization, P, induced by the radiation in the simplest case (of a single monochromatic wave propagating through a crystal of high symmetry) has the form

$$P = \chi^{(1)}E + \chi^{(2)}E^2 + \chi^{(3)}E^3 \ldots \tag{1}$$

The susceptibilities $\chi^{(1)}$, $\chi^{(2)}$, $\chi^{(3)}$ ... are related to the optical properties of the medium: as examples, the linear optical properties are related to $\chi^{(1)}$, the second harmonic generation is determined by the magnitude of $\chi^{(2)}$ and the intensity dependence of the refractive index is related to $\chi^{(3)}$. The science of nonlinear optics developed with the observation of second harmonic generation by Franken *et al.*[1] in 1961 and the theoretical prediction of a series of nonlinear effects by Bloembergen[2] and co-workers.

The large powers required in the early work made it seem unlikely that digital optical computing, using optical circuitry with nonlinear optical circuit elements, would ever be practical. It was nearly 20 years before the experimental discovery that a 30-mW continuous wave (CW) laser beam could be significantly self-defocused in a narrow gap semiconductor produced the surprising conclusion that third order nonlinearities ($\chi^{(3)}$) could have a value of the order of 10^9 times larger than the nonlinearities evisaged by Franken and Bloembergen[1,2].

To understand the significance of this giant nonlinearity to a possible future of optical computing, consider the physics of computation (see ref. 3). Information is essentially stored as energy and switching a device from logic 0 level to logic 1 level requires a definite switching energy. This energy must be greater than the thermal energy of the device kT (and will usually be several hundred times kT) and scales as the size of the device. There will usually also be a trade-off between speed and power. A further absolute requirement for electronic or optical logic is *the necessity of restoring* a logic level after each switching action so that errors resulting from imperfect devices and signals do not accumulate. Restoring logic must have power gain and this power is normally drawn from a power supply independent of the signal channel. (The requirement of power gain is obviously necessary, that is, one switching device must be capable of driving at least the next in the series so that a circuit containing multiple elements can be constructed.) These considerations appear to be applicable to all systems of signal processing and computing.

The response of semiconductor micro-electronics to the high data rate requirements of digital signal processing and computing has been to increase switching speeds and further miniaturize components in the form of very large-scale integration (VLSI). Transistors made from gallium arsenide have been reported with effective switching times as fast as 12 ps. However, this will not necessarily solve the problem of coping with high data rate as the processing time in conventional computers is many times the logic switching time because of the necessity of transferring information to the next part of the circuit. This involves capacitance time constant limits as pointed out, for example, in ref. 3: VLSI does not solve the time-constant problem because, as the length of a wire shrinks by a factor α and the cross-sectional area of the wire is reduced by a factor α^2, the capacitance C of the wire decreases by this factor α while the resistance R increases by the same amount. Thus, the time constant RC remains the same and the input charging time remains unaltered independent of scaling.

Turning to the computer as a whole, the standard method of communication in use today connects the logic unit with the memory through an address device. This reduces the number of interconnections but can only address one storage element at a time. This widely used scheme was first suggested by von Neumann but, rather than being given credit for this most practical innovation, he is now rather undeservedly blamed for this so-called "von Neumann bottleneck". The timing problems associated with circulating logic signals around a one-dimensional processor of this type ('clock skew') combine to indicate that future problems in digital computers are likely to be those of communication. This may apply at all levels (architectural, bus and chip) and stems from the use of time multiplex to compensate for the inability of electrical methods to communicate many channels of information in parallel.

Digital optics

Present practice has seen the invasion of electronically based communication by optical methods through the use of optical fibres in long-range telephone lines. The higher carrier frequency used gives potentially higher bandwidth, although electronic

limitations on modulation techniques have restricted our ability to exploit this greater information carrying capacity fully. Currently, long-range transmission made possible by low signal attenuation has been used. The use of optics for processing information has so far been handicapped by the absence of optical circuit elements. The optical methods have promising implications in areas that are currently difficult for existing technologies; these include image processing and recognition, sorting, radar array signal processing, machine vision and artificial intelligence.

During the past 20 years there has also been a body of work often called optical computing practised by optical researchers using linear processing devices, for example, the use of spatial filters and Fourier transform processes in image processing. Ideas include the use of symbolic substitution, residue arithmetic, vector-matrix multiplication and primitive processing experiments using a liquid crystal light valve as a hybrid (that is, optics with electronics) spatial light modulator but not itself capable of all-optical logic action. Most of the reports are theoretical proposals of what might be done (see ref. 4).

The major stumbling block in the development of digital optics has been the absence of nonlinear all-optical circuit elements, preferably capable of fabrication in the form of two-dimensional arrays and of small enough size to have small switching energies and high speeds.

The main purpose of this article is to review the latest state of research in all-optical nonlinear logic switches, amplifiers and memories and to show that, with the experimental realization of the first all-optical circuits, the time is now ripe for the combination of these two previously separate research efforts, optical bistability and linear optical computing techniques.

All-optical circuit elements

Recent progress, which has produced practical all-optical circuit elements, is based on the combination of optical nonlinearity and feedback. This has led to the concept of 'optical bistability' and hence to a whole family of devices based on a common set of physical and mathematical principles. The series of devices includes optical logic gates, bistable memories, amplifiers (sometimes termed optical transistors or transphasors) and power limiters (see ref. 5 for a recent Royal Society, London, conference). The meeting of the International Commission of Optics in Sapporo[6] presented six papers on optical computing and began the process of introducing the optical bistability community to the optical information processors. The recent Optical Society of America conference[7] at Lake Tahoe continued this process; a three-element loop processor, using all-optical interactions, was reported, together with experiments on parallel optical computing, using liquid crystal light valves as the nonlinear elements.

A seminal paper on optically bistable devices was that of Szöke *et al.*[8] who in 1969 proposed that a Fabry-Perot optical resonator containing a saturable absorber as its spacer layer could exhibit two states of transmission for the same input intensity. This simple bistable action was predicted to arise from the existence of a high internal optical field at constructive interference given that sufficient intensity had been incident on the resonator to bleach the absorber. To reach this condition required a greater input intensity than that required to maintain it. By contrast, at low input intensity the non-bleached absorption held the transmission of the device at a low level. In practice this condition is quite hard to achieve experimentally and the experiments described do not in fact show optical bistability. Observation of optical bistability was not made until 1976 when Gibbs *et al.*[9], using an interferometer containing sodium vapour, observed bistable transmission but deduced that the dominant mechanism was refractive, involving a shift in resonator frequency, rather than absorptive. Nevertheless, effective refractive nonlinearity resulted from a saturation of the atomic absorption. Such a device, although using only milliwatts of power, was relatively large (centimetres in length) and relatively slow (milliseconds) compared with electronic circuit components.

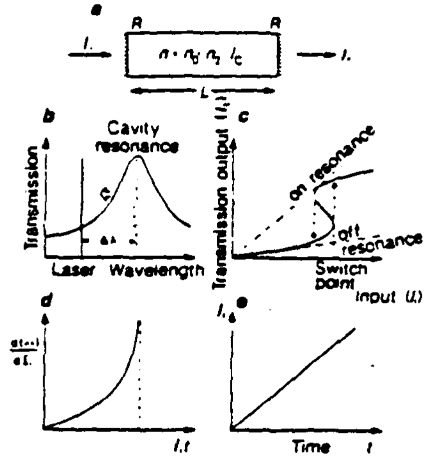


Fig. 1 The dynamics of switching in a Fabry-Perot etalon (see text for explanation of individual diagrams).

In the same year, my colleagues and I made the surprising discovery of giant nonliner refraction, defined as a function of the intensity, I, from

$$n = n_0 + n_2 I \qquad (2)$$

(where the nonlinear refractive index n_2 can be measured in units of cm^2 per kW) present in the narrow bandgap semiconductor InSb. The immediate implication was that a bistable resonator could be made of micrometre dimension and, as the effect was shown to be electronic, would be fast, probably on a nanosecond timescale. A second implication was that one beam could modulate the optical properties of a small slice of semiconductor and affect a second beam, thus making an optical modulator or optical transistor. The details of both the refractive and associated absorptive nonlinearities at milliwatt powers are described in ref. 10 but, retrospectively, the effects could be recognized in the earlier work on the spin-flip Raman laser. Both the device possibilities described above had been practically realized in InSb by 1979 (ref. 11) in which continuous wave laser beams were used, leading to steady-state operation and true optical bistability, as well as the observation of gain in an optical transistor[12].

Simultaneously and independently, optical bistability was reported by Gibbs *et al.*[13] using pulsed dye laser radiation in the larger gap semiconductor GaAs. Larger absorption and a smaller nonlinearity in this material, however, prevented steady-state operation and this observation was therefore quasi-dynamic and did not permit the demonstration of logic levels or differential amplification.

Origin of giant nonlinearities

The physical explanation of the large nonlinearities in both these semiconductors involves the excitation of electrons to give some degree of saturation or 'blocking'. In the case of InSb, exciton effects are negligible in the conditions of the experiment and a plausible explanation has been given by Miller *et al.*[14] in terms of a 'dynamic Burstein-Moss' shift of the band edge. Physically, a number of electrons ($\sim 10^{15}$-$10^{16}\ cm^{-3}$) are excited into lower conduction states by the laser photons. The incident intensity I generates an equilibrium number of electrons. Subsequent band filling, following thermalization, modifies the absorption edge and by application of the Kramers-Kronig relationship causes a change in refractive index,

$$\Delta n = \frac{\hbar c}{\pi} \int_0^\infty \frac{\Delta\alpha(\hbar\omega')}{(\hbar\omega')^2 - (\hbar\omega)^2} d(\hbar\omega') \qquad (3)$$

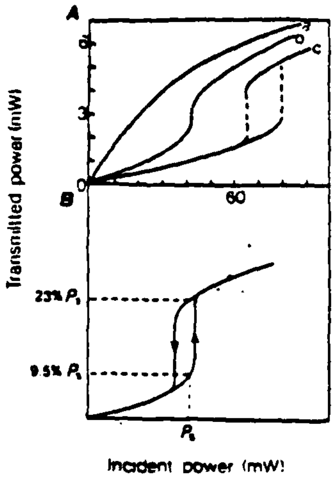


Fig. 2 *A*, Experimental observations of a family of characteristics of an interference filter bistable device illuminated by the argon ion laser line of wavelength 514 nm, obtained by changing the initial detuning from resonance of the etalon. The family of curves *a–c* shows the characteristics obtained for different values of the initial detuning parameter $\Delta\lambda$ set by the angle of incidence. If the initial condition is close to resonance, power limiting action is obtained (*a*) as the transmittance can only fall as intensity rises. As $\Delta\lambda$ is increased, curve *b* begins to kink, showing a greater change in output than for an incremental change in input, and thus exhibiting differential gain. This is responsible for optical transistor or transphasor action. Further increase of $\Delta\lambda$ leads to bistable loops of varying width (*c*). *B*, Experimental observation of optical bistability in the interference filter illuminated by light of wavelength 528 nm. The stability of the system is sufficient to allow operation to within ~4% of the power required to switch.

where $\alpha(\hbar\omega')$ is the interband absorption coefficient at photon energy $\hbar\omega'$. With $\Delta n/\Delta I = n_2$ we can obtain an analytical expression for n_2. This calculation gives a good description of the resonant refractive nonlinearity in InSb, InAs and CdHgTe.

Device physics

The simplest configuration which provides optical feedback is a simple Fabry–Perot etalon containing a nonlinear refractive material (Fig. 1*a*). Its optical thickness is given by

$$nL = (n_0 + n_2 I_c)L \tag{4}$$

(where L is the thickness of the material) and this changes with the internal intensity I_c. Consider now the transmission of such an interferometer as a function of wavelength (Fig. 1*b*): if we start in an initial condition where the illuminating laser wavelength is detuned from maximum transmission by a wavelength increment $\Delta\lambda$ (Fig. 1*b*), we see from Fig. 1*c* that the relation between output and input would give rise to, in the case of a linear device, a straight line of low slope: if the device were tuned to resonance and there were no absorption loss, output would be related to input by a line at 45°. If we now increase the intensity from the initial condition, the nonlinear resonator tends towards resonance as its optical thickness changes with intensity. This gives rise to a nonlinear relation between output and input. However, as we approach resonance, the internal field circulating within the resonator itself builds up according to

$$I_c = I_i T(\lambda)(1+R)/(1-R) \tag{5}$$

where I_i is the incident intensity, $T(\lambda)$ is the transmission as a function of wavelength (as in Fig. 1*b*) and R is the (constant) reflectivity of the resonator mirrors. Thus, at resonance the internal intensity is at its maximum where $T = 1$ and is amplified by the term $(1+R)/(1-R)$. This gives rise to positive feedback. As resonance is approached, the internal field builds up and the rate of approach to resonance depends on the change in optical thickness which itself depends on the magnitude of the internal field. The rate of approach to resonance thus speeds up. This can be readily expressed through the expression

$$\frac{d(\Delta\lambda)}{dI_i} = T(\lambda)\Big/\left(\frac{T_{max}}{2n_2 L} - I_i\frac{dT}{d\lambda}\right) \tag{6}$$

Figure 1*d* shows a plot of the rate of approach to resonance as a function of either incident intensity I_i (or as a function of time if we assume a linear ramp of I_i against time: Fig. 1*e*). As I_i increases, $dT/d\lambda$ also varies, the denominator of equation (6) tends to zero and the rate of approach to resonance becomes infinite, leading to rapid switching. This gives a physical feel for the dynamics of the switching process and leads to optical bistability (Fig. 1*c*), that is, two values of the transmission for one of the intensity. Two features are noteworthy: (1) a large value of finesse (ratio of spacing to half-width of the transmission peaks) makes it easier to obtain bistability; and (2) the form of the characteristic (output versus input) changes with the initial conditions, that is, it depends on $\Delta\lambda$.

Such a set of characteristics is illustrated in Fig. 2, giving a set of experimental results for nonlinear interference filters incorporating ZnSe films as the active layers (curves *a–c*). Thus, by a simple change of initial conditions a whole family of optical devices can be produced.

Figure of merit for optical devices

An analysis of the factors concerned in the design of such optical circuit elements has been given by Miller[15], who shows that the quantity which gives a figure of merit (in terms of nonlinear Fabry–Perot cavity optical switching in the presence of linear absorption, α) is $n_2/\lambda\alpha$. This determines the lowest critical value of input irradiance I_s for a device of given size to obtain bistable switching or nonlinear characteristic:

$$I_s = \frac{\alpha\lambda}{n_2} f(R, \alpha L) \tag{7}$$

The result is physically sensible as the switching power will be lower for a larger nonlinearity n_2, the shorter the wavelength the smaller will be the refractive change required to effect a change from constructive to destructive interference (that is, $\Delta(nL) = \lambda/2$) and for a smaller absorption (assumed linear in this analysis) the longer the device may be for a given loss. The function f optimizes cavity properties. Values of $n_2 = 0.1\ \mathrm{cm^2\,kW^{-1}}$ and $\alpha = 10\ \mathrm{cm^{-1}}$ give useful devices of thickness (L) 50–200 μm, bearing in mind that a fractional change $\Delta n/n \sim 10^{-3}$ is usually required.

To obtain favourable values for $n_2/\lambda\alpha$, we make particular use of effects resonant with the energy gap E_G in semiconducting materials. For electronic nonlinearities it can be shown that $n_2 \sim 1/E_G^3$ so that $I_s \sim 1/\lambda^2$. In addition, the considerations made earlier suggest that the smallest possible devices should be constructed. Diffraction limits suggest that the area limit will be $\sim(\lambda/n)^2$ and thus, although the nonlinearity is clearly large at longer wavelengths for small gap materials, the interference conditions and device size favour shorter wavelengths. Analysis of the detail of the frequency dependence of the nonlinearity, together with the attendant unwanted losses, has not so far indicated an optimum wavelength.

Semiconducting compounds have shown promising results: InSb with a typical working wavelength (5.5 μm) corresponding to 1,820 cm^{-1} is one material for which there are sufficient available laser frequencies to undertake a detailed examination of the frequency and hence resonant behaviour of n_2 and α near the band gap[5,14].

GaAs has also been investigated[6,13] but differs by having a strong exciton feature near the absorption edge. The nonlinearity $n_2 \sim 10^{-3}\ \mathrm{cm^2\,kW^{-1}}$ is quite practicable for devices, but the absorption coefficient in epitaxially grown material is so far giving values of $\alpha \sim 10^4\ \mathrm{cm^{-1}}$. Thus, thicknesses are restricted to a few micrometres and thermal stability poses a problem.

Both the above materials show negative values of n_2 caused by electronic effects. Carrier lifetimes are in the range of tens to hundreds of nanoseconds. The nonlinearities can be 'switched on' more quickly than this time interval by rapidly introducing carriers with relatively intense pulses. The relative figures of merit between GaAs and InSb favour InSb by a factor of 10^5, which has given InSb the advantage of steady-state operation but the disadvantage of a low operating temperature of 77 K, whereas GaAs has operated at room temperature and at shorter wavelengths. So far there are no reports of true steady-state operation for GaAs material. Other materials, which have been used in purely optical switching devices, are reported in ref. 5.

There is a second useful form of nonlinearity involving thermal shift of the band edge resulting from temperature rise of the bulk of the nonlinear material. This effect also resonates with the band edge and is associated with moderate values of absorption coefficient. The quantity taking the place of n_2 is an effective $n_2 = (dn/dT)\alpha(LL'/\kappa_s)$ where L, L' introduce dimensions of film and substrate thickness and κ_s is the thermal conductivity of the substrate. Interference thin film structures using ZnSe (ref. 16) have given promising results for refractive switching, and bulk ZnSe, CdS and GaAs also show 'optical bistability by increasing absorption' where the thermal shift introduces the required feedback[5].

Requirements for optical computing

Since our proposition is to undertake logic operation by means of these optical circuit elements, we can define some of the requirements if they are to be put together in the form of optical circuits to construct an optical computer. Requirements are as follows: (1) High contrast. A logic device needs to show a large change between logic 0 and logic 1 levels. (2) Steady-state bias. To make various different logic gates it is necessary to alter controllably optical bias levels. In terms of optical bistability this means that the device can be 'held' indefinitely at any point on the characteristic with a CW laser beam—the 'holding beam'—and implies a degree of thermal stability. Devices based on InSb at 77 K and on ZnSe at 300 K have been the first to show this behaviour. (3) External address. For logic functions it is clearly necessary that separate external signal beams can be combined with the holding beam to switch the device. The switching energy in fact is derived from the holding beam and this switched beam propagates in transmission or reflection as the output beam to the next devices in an optical circuit. (4) The elements must be cascadable. This means that the output of one device must be sufficient to switch at least one succeeding device. The ability to set a CW holding beam near to switch point in fact fulfils this condition because the extra increment is then small compared with the change in output even in the presence of loss. As each device has its own 'power supply' (that is, holding beam), logic levels are restored. (5) Fan-out and fan-in. The probable advantages of parallel processing in optical devices emphasize the ability that one device can drive many succeeding devices, probably using free space propagation for address. The summed effect of several elements can readily be focused on one device to achieve fan-in. (6) High gain. Items (4) and (5) demand that the elements show a value of differential gain ratio of change in output to change in input of >1. (7) Arrays. The technology shall ideally be such that two-dimensional arrays are easily constructed. (8) Power and speed. Low power per device is a necessity and this should preferably be of the order of milliwatts or less; this is aided by fabricating small-area devices. Speed and power will be interchangeable but speed itself will vary with use: for parallel arrays, microseconds will suffice and for one-dimensional circuits, subnanosecond or picosecond switching times are desirable.

Experimental results

Figure 3a shows an early result for an InSb resonator operating in transmission. The critical switching power is 20-25 mW incident on a diameter of ~200 μm. If the device is set up in

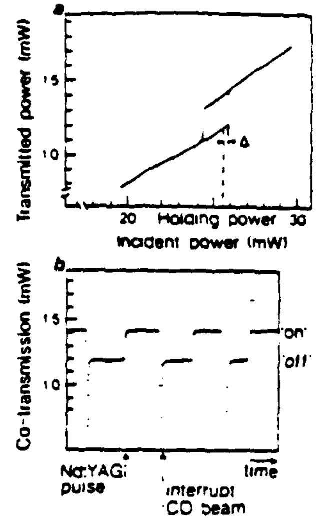


Fig. 3 a, Transmission characteristic of an InSb etalon illuminated by a steady-state carbon monoxide laser operating at a wavelength of 5 μm. The laser power is constant at a level slightly below that required to switch. b, The change in transmission induced by the absorption of 35 ps long pulses from a Nd:YAG laser (1.06 μm) and total energy 5 nJ.

this way with the holding beam adjusted in intensity to be just short of the switch point, the device may be externally addressed. Figure 3b shows the results obtained using a single 35-ps long pulse of 5 nJ energy from a Nd:YAG laser. The arrival of the single pulse is sufficient to trigger the switching and the device remains in the 'on' state. Interrupting the holding beam returns the switch to the 'off' state, and as can be seen the logic levels are extremely stable. The device was further developed to show explicit AND gate operation by dividing the switching pulse with a beam splitter and observing that the switching energy was quite definite so that the device could be set to switch only when both pulses were incident (in addition to the steady-state holding beam) and would not switch with a single pulse. The device is therefore acting as an AND gate and an optical memory.

We may deduce that sufficient free carriers were induced near the surface of the InSb resonator to cause sufficient change in optical thickness to initiate switching within ~3 ps. W. Kaiser (personal communication) has shown recently that this nonlinearity can be switched 'on' in <7 ps. Therefore, the probable switching time is limited only by the macroscopic effect of internal field build-up within the resonator. This depends only on the round-trip time for the 210-μm thick etalon and amounts to ~8 ps. We may infer that picosecond timescale optical logic has been demonstrated.

First digital optical circuits

With CW steady-state hold, an InSb logic device can be switched with incremental power as little as 3 μW (Fig. 3). Thus, the output change of the device operating in reflection (Fig. 4), ~4 mW, should be sufficient to switch a succeeding device. A first experiment using two optical itches, A and B, is illustrated in Fig. 4. This first example of an or .cal logic circuit is equivalent to an XNOR gate.

Thin-film interference structures comprising alternate layers of high- and low-index material deposited on a flat substrate such as float glass are a convenient way of constructing a resonator equivalent to the devices described earlier. If sufficient nonlinearity can be induced in the very thin interference layers the technology has many advantages, particularly that very uniform devices covering many square centimetres can be easily made. Thus, large-area arrays are readily possible. Typical total thicknesses of films for a device with 15 layers are only ~2-3 μm.

39

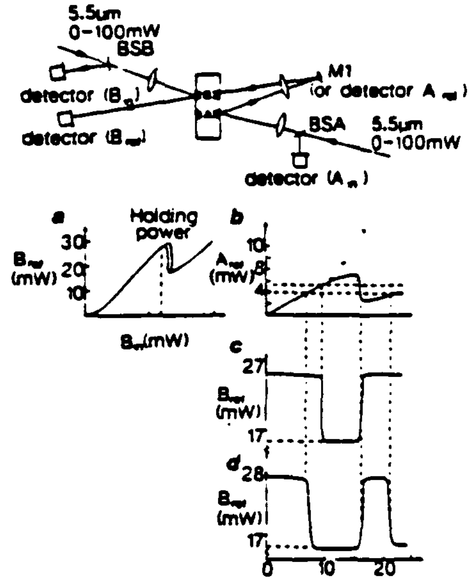


Fig. 4 The experimental arrangement used in the fabrication of an XNOR gate by the coupling of two InSb logic gates and the characteristics of the individual (*a*, *b*) and coupled (*c*, *d*) gates. Optical switch A is addressed by a holding beam which is reflected from the element and focused on to a second element, B, 500 μm away on the same crystal slice. This second element is held near to switch point by a second holding beam illuminating it from the opposite side. The output of A then acts as the address beam for B and the optical circuit output is observed as the reflection from B (B_{ref}). The figure also depicts the circuit inputs and outputs as the input to A (A_{in}) is steadily increased until a point is reached where its reflected output acting as address on B causes this second gate to switch to a lower state in reflection (*c*), (*d*). B remains in this lower state until A itself switches down in reflection, when B simultaneously switches up in reflection. B remains in the up state as the input to A is further increased until it again switches down when the reflected input from A is high enough. If A were to be held close to switch point and addressed with a pulse, information would be stored in A. If the holding beam on B is at the same time too far from switch point, the information will not be transferred from B. If this holding beam is then programmed to come within the range where B can be switched, the information can be transferred to B and by an appropriate reduction of the holding beam on B the information can be moved from A. The device can therefore be made to act as a shift register. With further holding beams defining optical circuit elements across the crystal, these forms of optical logic become indefinitely extensible.

The heat capacity of pixels of micrometre dimension is low and implies that incident power of ~10 mW can cause a temperature rise of 50 °C in ~1 ns. The important physics lies in the control of heat sinking by means of the conductivity of a relatively massive substrate. Figure 2 shows the very promising results obtained using ZnSe as the active film which has a band gap conveniently resonant with argon ion laser wavelengths at 514 and 528 nm. The important property recently found is that these logic switches are sufficiently stable to permit construction of optical circuits of a similar nature to those already demonstrated with InSb[5,17].

Optical computers

I have described the contemporary performance of several optical circuit elements and shown that, in principle, all logic functions familiar in electronics can be reproduced by these nonlinear optical devices. Therefore, it should be possible to construct an all-optical computer.

There exists a considerable body of literature reporting the research of a number of groups on the subject of optical comput-

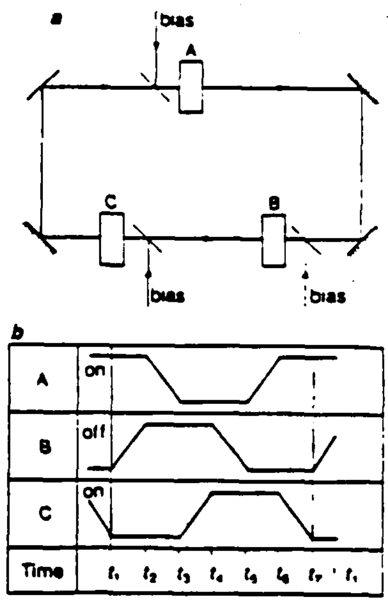


Fig. 5 *a*, A schematic model of the arrangement used to demonstrate a three-gate optical delay which forms the basis of a loop processor. *b*, The clocking of the individual bias levels to the three gates.

ing[4]. To date this has not included the use of the nonlinear elements discussed here. Nevertheless, the advantages of using optical methods for various computations have been apparent for several years. Historically, most optical computing systems have been of analog type in which information is stored and processed as a continuum of signal levels. There are major drawbacks to such analog systems, including limited flexibility, noise accumulation and input/output device limitations. The advantages of optical binary logic, now in principle possible with the devices described here, may be crucial in progressing optical computing into a new practical technology. One would wish to avoid as many photon–electron or electron–photon conversions as possible. The considerations can be made at several levels, including individual optical logic devices (gates and arrays of devices) as the first level. The second level considers optical communications interconnections and input/output among the logic gates, among arrays and among circuit boards or processors. The third level considers the possibilities of new computer architecture to take advantage of the inherent parallelism of optics.

Thus, one might see the advantages of optical methods in two separate ways: (1) Speed of switching and communication. I have shown that individual optical logic operations can almost certainly be performed on a picosecond timescale. Combining this with the output characteristics of CW mode-locked lasers, it is possible to conceive bit rates approaching frequencies of THz for an exact analogue of electronic digital logic. It is not yet clear that recovery times can be imposed on the optical logic switches at acceptable power levels to ensure continuous operation. The possibility exists of series-to-parallel and parallel-to-series conversion on picosecond timescales. Fast, high bandwidth communication will also be consistently available. Optical methods may defeat clock skew. (2) Parallelism and communication. The logic devices described here are at this time the only ones which have been capable of providing steady-state information holding and cascadable operation. They have time constants in the range 100 ns–100 μs. This suggests that the first approach to the use of optics should emphasize the use of parallelism and accept cycle r[illegible] similar to existing electronic speeds. The intriguing intellectual challenge is to use the flexibility of optics to provide appropriate interconnections:

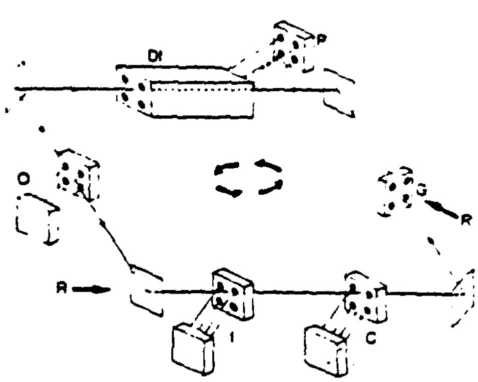


Fig. 6 The types of optical computer being proposed use the parallelism of optics by using arrays of elements. The iteration present in the type of problem which could be solved on such a computer is achieved through the use of a loop processor architecture. The individual arrays of elements may have to be flexible in their uses in the processor, identical arrays may be used as input (I), clock (C), gain (G), program (P), dynamic interconnect (DI) and output (O). Each array may require a separate power supply refresh (R).

requirements could be to shift a logic array pattern by one element on each cycle or to achieve a 'perfect shuffle'. Other schemes include a vector-matrix multiplication involving a fan-in from a matrix array to a column followed by rotation of this column between output and input for successive cycles.

In my laboratory a number of optical-computer architectural components have been designed including memory units, clocks, a programmable processing stage and a simple full adder using simultaneously the transmission and reflection from an optical gate[18]. The necessity to store the calculation before communicating to the next element is important in the optical case: propagation at the speed of light otherwise would mean that all elements would be simultaneously addressed (admittedly, avoiding clock skew[3]). Delays can be readily implemented by suitably programming the holding beams so that the optical bias is only set to receive a signal when the device is clear of other signals. As a first step, consistent with the elements described, this could be achieved electronically making an interesting interface to existing computer technology. At the time of writing, a three-element loop processor with optical bias delay loop clocking has been implemented. Acousto-optic modulators, controlled by a microcomputer, provide both optical bias levels and input data pulses (Fig. 5). In principle, indefinitely extensible optical logic has been demonstrated. A generalized optical computer using these ideas may take the form shown in Fig. 6.

Conclusion

One may conclude that practical optical circuit elements have been demonstrated, that all binary logical operations can be performed and that the first digital optical circuits have been shown to be practical.

At this time the most favourable method of applying optical methods to digital optical computing seems to be the use of digital arrays of gates speaking (in an optical sense) to further arrays of gates with the possibility of fixed or variable intercommunications, fan-out and fan-in using free space propagation as efficient ways of 'optical wiring'. It remains to be shown that holding powers and switching energies are sufficiently small to allow operation with existing lasers and devices. Scaling from present experimental results, one may deduce that, for a visible wavelength device, a single logic element area of $\sim 1\ \mu m^2$ should be practical, which, with a 250-ns switching time, should require 0.2 mW of holding power. Perhaps 10% of this would be absorbed. The 10 W available in the 514-nm argon ion line from a typical commercial laser would allow the operation of 5×10^4 gates assuming perfect optics. This would give the array a rate of information processing equivalent to 2×10^{11} logic operations per second assuming that all the gates are used simultaneously. A similar number of logic operations could be achieved by using small-gap semiconductors in the infrared where the electronic processes would allow faster switching but the packing density and device size would be greater. The optimization of the existing devices and the use of different materials may allow the synthesis of arrays capable of much higher data rates.

Even allowing for imperfections, which would reduce these numbers, the data rates are sufficiently high to encourage the exploration of their use for tasks where conventional one-dimensional sequential digital electronics finds difficulties, such as pattern recognition, artificial intelligence, sorting and specialized computational problems. Equally interesting will be applications such as power limiters, optical noise reduction and laser projectors and displays.

International development programmes

Several collaborating groups are actively involved in research in these areas. The principal effort in Europe is the European Joint Optical Bistability Program, a multinational project on optical logic circuitry and the basic physics of optical bistability involving eight universities and institutes in Britain, Belgium, West Germany, Italy and France. There are a further eight European laboratories associated with the programme. The research support scheme was established in 1984 by the Commission of the European Communities, through its Committee for the European Development of Science and Technology.

In the United States there are two main collaborative groups: The Optical Circuitry Cooperative is currently being established by the University of Arizona, which is involved with over 10 American companies with an interest in optical circuitry. The five-year programme is funded by these companies, the university and a NSF grant. The Pentagon recently announced the formation of a consortium of nine research organizations, including seven universities which will be funded as part of the Strategic Defense Initiative to produce an optical computer.

Japan's Optical Computer Group was formed in 1983 with the purpose of providing the opportunity for its 75 members to exchange information and ideas in the field of optical computing. The group has four to six meetings every year and publishes a newsletter (*Opcom news*) between meetings. The group is associated with the Optics Division of the Japanese Society of Applied Physics with members from Japanese universities and companies.

I thank Dr F. A. P. Tooley for help in preparing this manuscript as well as all members of the Heriot-Watt optical circuits research group and the participating laboratories of the European Joint Optical Bistability project of the Commission of the European Communities.

1. Franken, P. A., Hill, A. E., Peters, C. W. & Weinreich, G. *Phys. Rev. Lett.* 7, 118–120 (1961).
2. Bloembergen, N. *Nonlinear Optics* (Benjamin, Menlo Park, 1965).
3. Mead, C. & Conway, L. *An Introduction to VLSI Systems* (Addison-Wesley, London, 1980).
4. *Opt. Engng* 24, no. 1 ... (1985).
5. Wherrett, B. S. & Smith, S. D. (eds) *Optical Bistability, Dynamical Nonlinearity and Photonic Logic* (Royal Society, London, 1984) and *Phil. Trans. R. Soc.* A313, 191–451 (1984).
6. Meet. Int. Comm. Opt., Sapporo (August 1984).
7. Conf. opt. Soc. Am., Lake Tahoe (March 1984).
8. Szoke, A., Daneu, V., Goldhar, J. & Kurnit, N. A. *Appl. Phys. Lett.* 15, 376–379 (1969).
9. Gibbs, H. M., McCall, S. L. & Venkatesan, T. N. C. *Phys. Rev. Lett.* 36, 1135–1138 (1976).
10. Miller, D. A. B., Mozolowski, M. H., Miller, A. & Smith, S. D. *Opt. Commun.* 27, 133–136 (1978).
11. Miller, D. A. B., Smith, S. D. & Johnston, A. *Appl. Phys. Lett.* 35, 658–660 (1979).
12. Miller, D. A. B. & Smith, S. D. *Opt. Commun.* 31, 101–104 (1979).
13. Gibbs, H. M. *et al. Appl. Phys. Lett.* 35, 451–453 (1979).
14. Miller, D. A. B., Seaton, C. T., Prise, M. E. & Smith, S. D. *Phys. Rev. Lett.* 47, 197–200 (1981).
15. Miller, D. A. B. *IEEE J. quant. Electron.* QE-17 306–311 (1981).
16. Smith, S. D. *et al. Opt. Commun.* 51, 357–362 (1984).
17. Smith, S. D. *et al. Opt. Engng* 24(4) (July/August 1984).
18. Wherrett, B. S. *Appl. Opt.* (in the press).

41

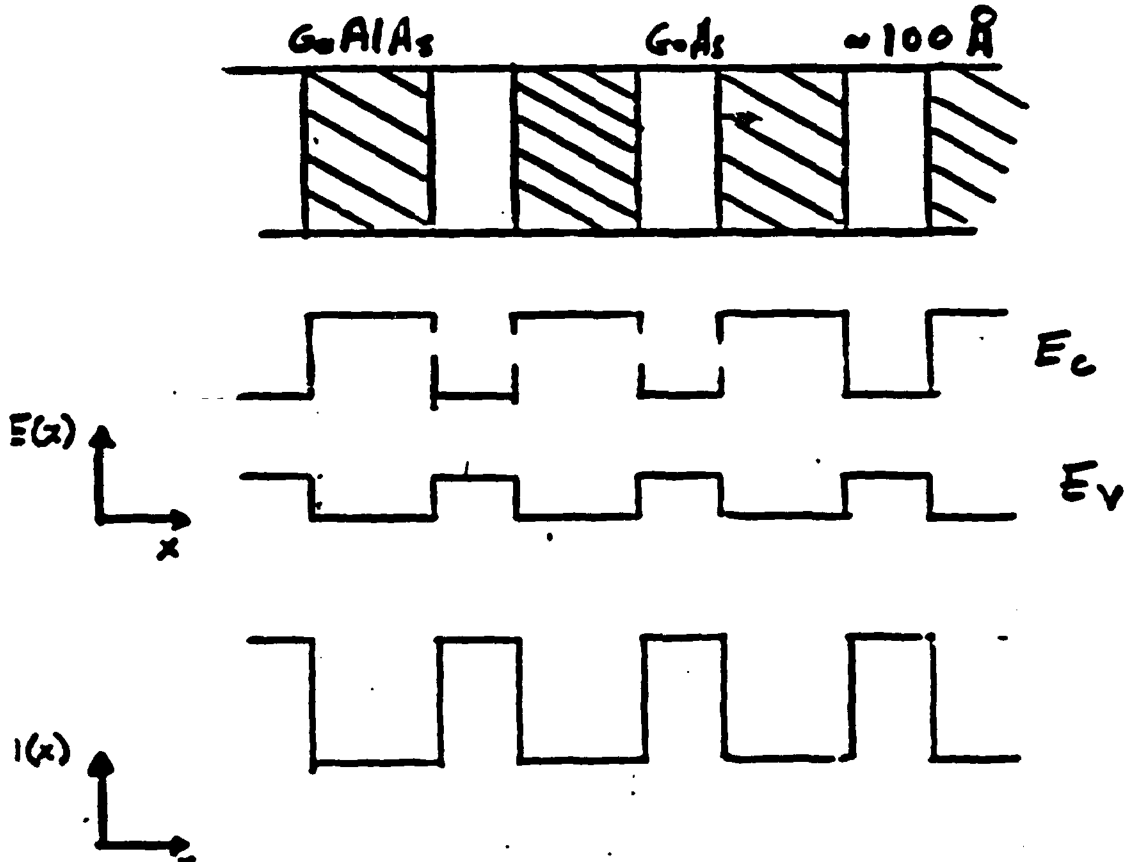
Synthetic Nonlinear Media

M. Fejer and R. Byer

Applied Physics Department
Edward L. Ginzton Laboratory
Stanford University
Stanford, CA 94305

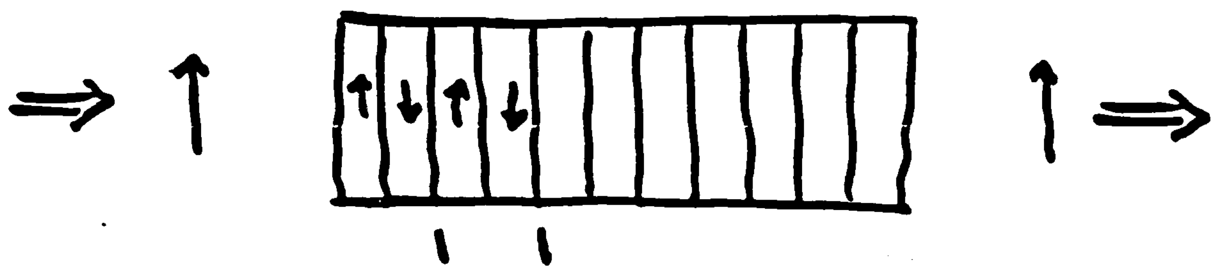
Multiple Quantum Wells

- Alternating layers of GaAs – GaAlAs
  – typically 100 Å spacing
- Spatially varying bandgap forms potential wells
- Stratified variations in index of refraction
- Piggyback on integrated circuit technology



Periodic Poled

$LiNbO_3$



6.8 microns period

$d_{eff} = \left(\frac{2}{\pi}\right) d_{33} \qquad d_{33} = 7 d_{31}$

$\eta_{SHG} = 17\times$ bulk Xtal

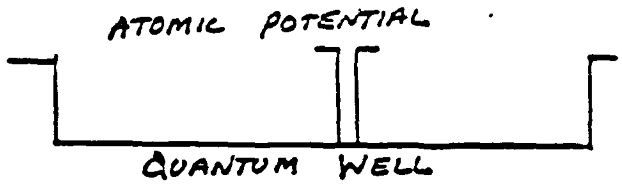
$= 7\%/\mathrm{W\,cm}$

Tailored Nonlinear Media

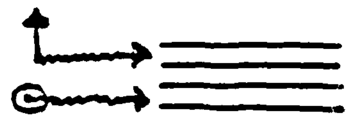
- Molecular
  - organics
  - beta-barium borate
- Microstructures
  - multiple quantum wells
- Macrostructures
  - waveguides
  - periodic media
- Combination of techniques attractive

Optical Properties

- Highly nonlinear
  - 10–100 x GaAs



- Birefringent
  - $\Delta n = 0.05$

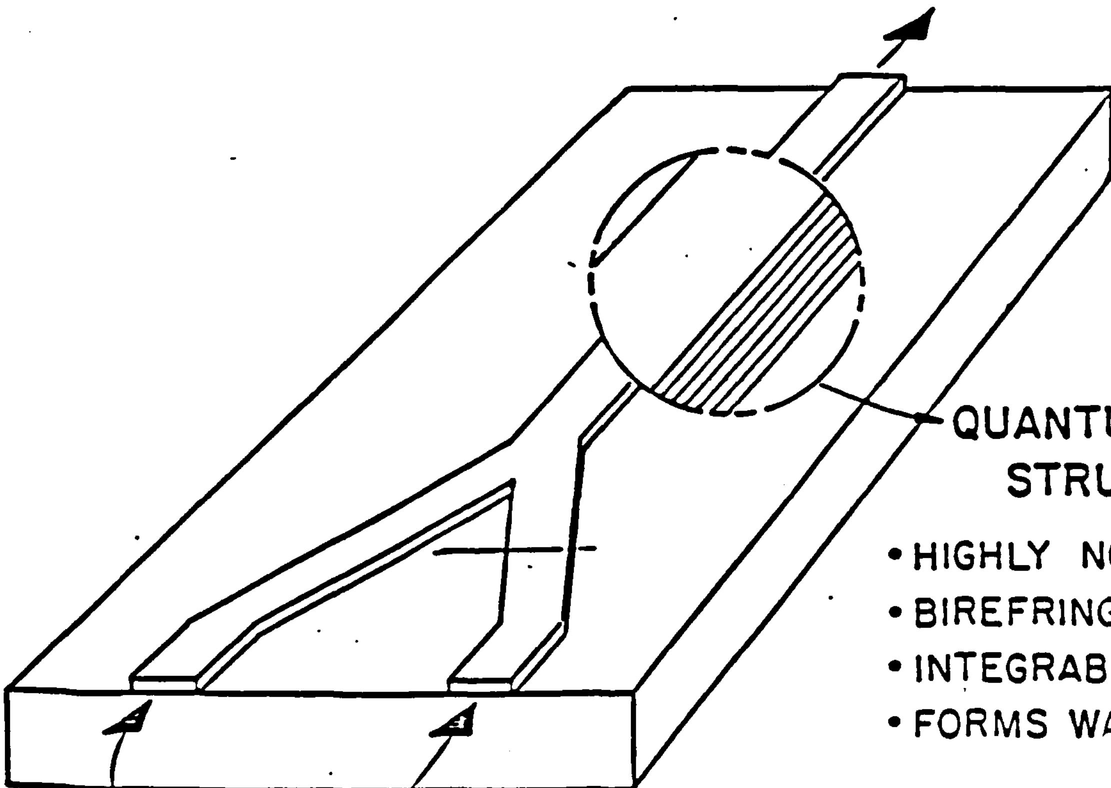


- Optical Waveguide
  - dispersion for phasematching
  - beam confinement for efficiency

NONLINEAR OPTICS IN QUANTUM WELL STRUCTURES

OUTPUT

- FREQUENCY DOUBLED
- MIXED
- MODULATED



QUANTUM WELL STRUCTURE

- HIGHLY NONLINEAR
- BIREFRINGENT
- INTEGRABLE
- FORMS WAVEGUIDE

INPUTS

- INTEGRAL DIODE LASERS
- OFFCHIP OPTICAL DEVICES
- MICROWAVE SIGNALS

Robert L. Byer
Professor Applied Physics
Edward L. Ginzton Laboratory
Stanford University
Stanford CA 94305
(415) 497-0226

November 6, 1985

CONCLUSION

NATURAL CRYSTALS

IMPROVED CRYSTALS

SINGLE CRYSTAL FIBERS

SYNTHETIC NONLINEAR CRYSTALS

OPTICAL | ELECTRICAL

OPTO-ELECTRONIC
MATERIALS

891 Allardice Way
Stanford, CA 94305
(415) 493-6192

4

E. "DoD Applications of Electromagnetic Launch (EML) Technologies" LJI-R-85-351 [W. Weldon, A. T. Atkin, S. Nozette, and B. Tapley (University of Texas, Austin)]

This white paper on "DoD Applications of Electromagnetic Launch (EML) Technologies" speaks to several potentially important application areas for this developing technology. These are (1) hypervelocity impact research which can run the gamut of materials research from equations of state, materials properties and processing, to vulnerability, and armor-antiarmor studies among others; (2) launchers and accelerators for spacecraft or aircraft, and (3) spacecraft propulsion. These ideas originate with a University of Texas group, part of which is already funded by DARPA for construction of a rail-gun facility. The latter's goal is to achieve 50 km/sec using a 1 gram mass. At lower velocities (20-30 km/sec) larger masses could be accelerated. Conducting materials, enclosed in an insulating sabot, could also be used, with a mass penalty. Current funding allows for measurement of projectile speed and integrity but not for impact studies. The last two topics represent relatively uncharted areas; the first could have nearer-term payoffs with information being generated at a rep-rate of ~1-2 shots/day. One might imagine an eventual national user facility for the impact studies.

F. Strategic Computing Applications Program LJI-R-85-348
[J. Boris (NRL) and P. Hammerling (LJI)]

This report describes a Strategic Computing Applications program which would apply new parallel processing architectures together with a novel algorithm of Dr. Boris to the surveillance, correlation, and tracking of multiple targets of interest. The ultimate goal of the program is the production of modules that could be carried on board a platform: ship, aircraft, or land vehicle. Another, nearer term, application would be as a simulator for battle management engagement scenarios. The battle management problem has been characterized by the recent Eastport Study Group as "the paramount strategic defense problem."

A critical element is the monotonic logical grid (MLG) algorithm alluded to above. This technique organizes the geometric information that descibes randomly-moving nodes in a way that ensures that near neighbors in space are also automatically near neighbors in the data arrays. As discussed in detail in the report, the MLG reduces the combinatorial problem from an order N^2 problem to an order N problem. The MLG algorithm also organizes the computational problem in a way that is ideally suited to the TMC "Connection Machine" (CM).

It is this combination of algorithm and architecture that should allow a *"quantum jump"* of a factor of $\sim 10^3$ (10^2 from the MLG and 10 from the CM) in the number of objects that can be processed per unit time.

G. "Four Papers on Robotics: 1. White Paper for the Development of the 'Super Robot', 2. Next Generation of Technology for Robotics, 3. Summary Report on Pipelined Computation of Dynamic Modeling Matrice for Serial Robotic Manipulators, 4. An Assessment of the Development and Application Potential for Robots to Support Space Station Operations," LJI-R-85-349 [D. Tesar, (University of Texas, Austin)]

These four, linked papers outline a large-scale program for robotics research which could, if implemented, have a large impact on DoD as well as civil use of robots.

The super robot represents the full integration of the most advanced computer technology (the super computer) with the most general mechanical architecture (serial, parallel, modular, etc.) to demonstrate an electronically-rigid system (similar to our latest fly-by-wire aircraft) capable of rejecting process disturbances in real time while producing high value-added products on demand. Today, high value-added operations are achieved primarily through the use of expensive, specialized and dedicated machines such as N.C. machines, automatic-screw machines, wire-bonding equipment for microcircuits, etc., where the robot performs the low-valued functions of handling of parts between these dedicated machines. By contrast, the super robot would be a fully-integrated and self-contained generic machine system capable of performing a wide spectrum of precision light machining operations completely programmable by the designer of the product (shoes, clothes, appliances, etc.) and fully responsive to the individual demands of the marketplace. This vision of robotics by Issac Asimov is the heart of the factory of the future, yet it not only does not exist, technical resources to make it possible are either in short supply or have not been

concentrated in a sufficient critical mass of expertise to make it happen.

Beyond the factory of the future there are applications of robotics to functions which involve hazards to humans such as space operations, operations on the ocean floor, ammunition handling under chemical or biological attack, processing of dangerous materials such as galium arsenide for advanced microcircuit technology, nuclear reactor maintenance, etc. In addition, special applications of real value to society, such as microsurgery, have yet to be dealt with even in the research environment. The concept of the super robot being proposed here would lay the foundation to demonstrate a science of intelligent machines sufficiently general to treat all of these devices and rewarding applications.

Much of the mechanical design philosophy in the United States derives from a period during which farm machinery, power plants, construction machinery, automobiles, airplanes, jet engines, etc. were brought to a high level of development. Much of this design is performed in terms of compartmentalized rules (the basis of an art and the opposite of a science) which are based on negative criteria (noise, wear, fatigue, instability, vibrations, mean time between failures, etc.). On the other hand, the factory of the future demands the use of operational criteria associated with the quality of the product of the machine which implies precision (rarely dealt with as a first priority in the academic world). The positive criteria of precision involves the control of the output of the machine to specified tolerances regardless of the

disturbances generated by the operation. To date not a single robot operates in terms of a real time dynamic model based on an accurate description of its system parameters in order to reject disturbances (i.e., the concept of closed-loop operation found in fly-by-wire aircraft). Furthermore, the negative design criteria of failure in the operation of large machine structures of the past (textile machinery, battlefield materiel, etc.) have little to offer for the design of precision microprocessing equipment of the scale suitable to microsurgery or microcircuits. Hence, relative to the level of technical integration required to meet future needs, no balanced science of intelligent machines is being developed.

Today, the drive to establish the factory of the future has led to vigorous development activity associated with CAD/CAM. Unfortunately, almost all of this activity is centered on the use of a collection of dedicated machines, each capable of a limited number of distinct critical precision functions which must be sequenced to create the finished product. On the other hand, the fully-integrated, self-contained, intelligent machine which is capable of producing broad classes of quality products fully responsive to the individual consumer does not exist in any form. In fact, the use of 15,000 robots in the U.S. at this time implies a penetration into the manufacturing workplace of not more than 1 in 1000 showing that robot implementation is far below the level needed to have real impact. This low level of penetration may be due partially to the fact that each of our major firms (IBM, GE, GM, Westinghouse) made one robot and then decided to purchase

robots from outside vendors or to buy subsidiaries either in Japan, Europe, or in the U.S. By contrasts, in Japan, each of the major manufacturing firms (Hitachi, Mitsubishi, Fujitsu, etc.) make their own robots. The contention here is that U.S. firms do not have the necessary in-house, balanced technical manpower to remain competitive in this leading-edge technology and are leaving it to their economic competitors. This lack of response to the threat of the trade deficit, exceeding $100 billion in value-added products, is at the heart of the present proposal. The goal is to employ existing component technologies (the super computer, computer vision, digital control theory), enhance emerging technologies (expert systems, artificial intelligence, metrology, mechanical architecture, computer architecture, CAD/CAM), and fully integrate them by means of a balanced science for intelligent machines. The super robot would be the most aggressive demonstration of this objective.

The simplest robotic architecture is a 6 degrees-of-freedom (DOF) serial system (one link, one joint, one link, etc.). To date two basic geometries have emerged. One is a structure similar to a coordinate axis (X-Y-Z) machine and the other is similar to a human arm. These simple structures are used because they represent very few design parameters and are designed primarily by intuitive means. The general 6-DOF serial robot system is described by 18 geometric, 36 mass, 36 deformation, and 18 control parameters (a total of 108) and represents a design complexity far beyond the means of existing expertise in industry. Beyond the serial structure, there are parallel structures

(walking machines with 4 or more legs), redundant structures to form systems from building blocks the way we now create computer systems, etc. What this means is that the design techniques for most future robotic systems do not yet exist and can only be developed by a very aggressive research program.

Similarly, no industrial robot operates in terms of a real-time, dynamic-model description to close the loop relative to the process it is performing which may generate significant disturbances in the system. This means that precision, light-machining operations such as drilling, routing, milling, etc., cannot be performed by reasonably-sized generic robots to the level of precision required. Disturbances due to forces equivalent to the specified load capacity of these robots can easily cause a deflection 20 times as great as the error represented by its repeatability (i.e., a 20 to 1 robot). The goal must be to measure these disturbances and to compensate for the resulting deformations (in order to maintain the desired level of precision) by means of a complete dynamic model evaluated in less than 10 msec (real time) by using the most modern computational hardware and software. This class of control would be equivalent to feed forward compensation (a technique now found in the very best Japanese Hi-Fi equipment) and is what is meant by an electronically rigid-robot system.

Over the past several decades, the electrical research community has made major strides forward in its technical depth especially enhanced by strong "pulls" from the civil and defense sectors. By contrast, mechanical technology has not kept pace

such that it is now perceived as a weak partner. Unfortunately, the mission objective of intelligent machines will require a marriage of these technologies as equals. Hence, in order to satisfy the super robot development mission, it will be essential to <u>create a fully-integrated science of intelligent machines based on a balanced development of all required electrical and mechanical component technologies</u>.

The concept of the super robot is the full implementation of a cohesive analytical description of generalized mechanical architecture with a major emphasis on the use of the super computer to benchmark the complete controlling equations for deformation, dynamics, adaptive control, and feedforward compensation for the effects of external or process generated disturbances in real time operation. The objective is to obtain results which are able to describe the operation of any general robotic structure, thus allowing for specialization to a given device suited to a unique application. This top-down approach (similar to the approach used to validate the development of the super computer itself) is completely missing in the development of robotics to date. Literally hundreds of design parameters are involved yielding potentially billions of possible systems. The optimal design and, therefore operation, of these systems is essentially unreachable with present small-scale bottom up technologies. That is why most industrial robots look either like an X-Y-Z measuring machine or a human arm, both of which are several orders of magnitude simpler than the general mechanical architecture. Consequently, the super robot effort is intended to

integrate all the previous analytical research of the team (and that of others) into a fully-operational parallel, modular, or mixed robotic structure.

The research program will concentrate on the use of super computer to dramatically accelerate the development of a science of intelligence machines because of its superior computational capacity to treat the full parametric description of a much more general class of robot structures. For example, the massive computational resources of the super computer make it possible for the researcher to think much more openly and freely of generic, top-down design and control strategies which should lead to a maximum level of productivity of new ideas and technology evaluated by complete simulations.

H. "Strategy for Complex Organization Modeling, Planning and Experiment (SCOMPLEX)," LJI-R-85-351 (B. J. West (La Jolla Institute)

This report outlines and motivates a mechanism for managing complex, large scale, scientific and technological research programs. This strategy (S) for complex (C) organizational (O) modelling (M), planning (PL) and experimenting (EX) (SCOMPLEX) is discussed with reference to examples drawn from physical processes relevant to DARPA programs. Drawing from our experience in the analysis of complex systems, we argue that: if the systems consist of a number of only weakly-interacting components, (example, ASW), then the traditional managerial schemes would appear to be adequate. If however the system consists of a large number of strongly-interacting or interdependent components (example, wave propagation in random media), then a new managerial scheme is proposed. This new scheme has the following elements:

1. A principal investigator as the Program Director and who has the final decision-making responsibility in all program areas.
2. An Overseer Committee consisting of the senior scientists in the program which will oversee the allocation of funds, personnel, etc. as well as the overall research direction of the program.
3. An External Scientific Advisory Panel (ESAP), whose members are chosen from the academic, industrial and government scientific committees to provide the best available advice and guidance from the outside experts in the field of programmatic interest.

4. A Technical Transfer Panel (TTP), whose members will be chosen from governmental and industrial settings to provide the best available advice and guidance in the technological implications of the scientific research, particularly for the services. It is suggested that the chair of this committee be the DARPA contract monitor for the program.
5. These four elements are duplicated at the level of the research conducted by each of the senior scientists in the program. In this way they each become the director of their individual project for which there is an Overseer Committee, an ESAP and a TTP. The managerial structure is thus seen to be self-similar and the number of self-similar levels depends on the degree of complexity of the program.
6. At each level of the proposed hierarchy the four elements are used in a self-assessment mode to determine if the research goals of the program are being realized and if not, what is required to realize their goal. They monitor, critique and guide the scientific research in a coordinated manner.

ADDENDA

LJI-R-85-351

DOD APPLICATIONS OF ELECTROMAGNETIC LAUNCH (EML) TECHNOLOGIES

BY

WILLIAM WELDON*, A. JAMES ATKIN*, STEWART NOZETTE†, AND BYRON TAPLEY†

NOVEMBER 1985

LA JOLLA INSTITUTE
P. O. BOX 1434
LA JOLLA, CA 92038
(619) 587-6000

THIS RESEARCH WAS SPONSORED BY THE
DEFENSE ADVANCED RESEARCH PROJECTS AGENCY
UNDER ARPA ORDER NO.: 3710
CONTRACT NO.: MDA903-85-C-0187

*Center for Electromechanics, University of Texas, Austin
†Center for Space Research, University of Texas, Austin

TABLE OF CONTENTS

Introduction

Electromagnetic launch (EML) technologies have been advanced by DOD research, including efforts supported under the Strategic Defense Initiative (SDI). This white paper provides an overview of potential DOD application for these technologies in near term research, development, and testing, which can compliment the applications in tactical and space based weapon systems. Three areas of potential interest are: 1) Hypervelocity impact research, 2) Launchers and accelerators for aircraft and spacecraft, and 3) Space propulsion.

Electromagnetic Launch Technology

EML technology can be generally divided into two approaches: the simplest being the railgun, the more complex and potentially more efficient being the coilgun. The electromagnetic railgun (Figure 1) consists of two parallel, metal rails separated by a distance typically equal to their width. A projectile with a conducting armature is placed between the rails at the breech of the railgun. If the breech of the railgun is then connected to an appropriate source of electrical current, the current will flow down one rail, across the armature and back up the other rail. The current flowing in the rails creates a magnetic

field between the rails, and the current flowing in the armature interacts with this magnetic field to produce a force which accelerates the armature and projectile down the gun barrel. This accelerating force, known as the Lorentz force, reaches levels of interest only at extremely high currents ($>10^5$ A). Unlike a thermodynamic gun in which the acceleration falls off as the hot gas expands, the acceleration in an EML can be held constant as long as constant current is maintained in the gun. Being the simplest of the EML's, the railgun has enjoyed the most rapid development. In recent years, masses as high as 300 grams have been accelerated to velocities in excess of 4 km/s while smaller masses (1-5 g) have been accelerated to velocities of 8-10 km/s. For comparison, conventional guns are practically limited to velocities of 1.5-2 km/s.

Coilguns, while being more complicated than railguns, also offer the promise of higher efficiency and greater control over acceleration. The basic concept for a coilgun (Figure 2) involves a series of stationary (stator) coils and a moving (armature) coil attached to the projectile. As the armature coil passes through each stator coil, current is directed into the stator coil so that the armature coil is repelled down the gun barrel. A variety of coilgun configurations have been considered which differ in the way current is supplied to the stator and armature coils. The coilgun is attractive since it does not require contact between the armature and stator, as the railgun does, and since its higher impedance leads to higher efficiency as mentioned previously. The additional complexity of the coilgun stems from the more complicated construction of the stator and the need to synchronize the current feed to the individual stator coils with the position of the armature.

Pulsed Power Technology

EML's of interest require electrical power of 100's of MW to several GW during launch. Although the basic operating principles of EML's have been known since the early part of this century, their enormous power needs kept them from being realizable until recent advances in pulsed power technology (PPT). PPT uses energy storage techniques to store energy slowly at moderate power levels and then deliver that stored energy in a brief, intense burst of electrical power. Energy may be stored electrostatically in capacitors, electromagnetically in inductors, electrochemically in batteries, or in the inertia of spinning flywheels Recent developments at UT's Center for Electromechanics (CEM-UT) involving the incorporation of specialized rotating electrical generator technology with inertial energy storage have made compact, inexpensive, portable pulsed power supplies available for driving EML's. Of course, the pulsed power supply must do more than just store energy. It must deliver the desired current at the appropriate voltage level in exactly the proper time frame. Two CEM-UT developed power supplies are capable of performing this crucial task for a variety of EML's as well as other applications. The first of these is the pulsed homopolar generator (HPG). Although the basic concept is over 150 years old, a portable HPG pulsed power supply has only recently become practical. Figure 3-a shows the principle of HPG operation. As a monolithic conducting rotor (flywheel) spins in an axial magnetic field a voltage is generated between the shaft and outer periphery of the rotor. If sliding contacts are applied to the shaft and rotor periphery, the generated voltage can be utilized to drive a current in an external circuit. As electrical energy is extracted from the HPG, the rotor slows because its inertial energy is being converted to electrical

output. Figure 3-b shows a compact, portable pulsed HPG developed by CEM-UT. This machine stores 6.2 MJ of energy inertially and can produce output currents up to 1.5 MA, stopping its 1,600-lb rotor from 6,200 rpm in a fraction of a second. CEM-UT developed HPG technology has been licensed to Parker Kinetic Designs, Inc. of Austin, Texas.

A second CEM-UT developed pulsed power supply, the compulsator, was invented in 1978. Whereas the HPG produces a single output pulse as it slows, the compulsator produces a burst or a continuous chain of pulses. Figure 4 shows the rotor and stator of a compulsator under construction at CEM-UT. This machine, which is designed to power an electromagnetic machine gun, will produce a burst of ten 2.5-kV, 1-MA, 2-ms pulses in one-sixth of a second.

Experimental Hypervelocity Impact Research

Experimental hypervelocity impact research can be greatly advanced by EML technologies. The current research tool available for hypervelocity impact research is the light gas gun (see Figure 5). As with other types of guns, the maximum velocity achieved is limited by thermodynamic properties of expanding gases: Light gas guns are limited to a maximum of 10 km/sec with macroscopic projectiles. Experimental results above this velocity have been obtained with projectiles in the size range of 10's of microns using plasma acceleration. In theory EML technology could overcome these limitations as a railgun could substitute for the light gas gun. The availability of a hypervelocity impact research facility which can explore the 10-20 km/sec domain could have many applications. In the space environment hazardous objects (both natural and man-made) with these

velocities occur. The existing data base regarding effects could be expanded so as to design better shielding against these threats. The survivability of space based systems is a key issue that must be resolved if active weapons systems are to be deployed. A large amount of hypervelocity impact research has been carried out during the Apollo era, and by DOD for its purposes. Much of this data and experience is still applicable. However, the potential of active weapons systems being deployed in space, and the problem of man-made debris in Earth orbit raises new questions that can only be addressed by actual experimentation.

An EML based facility can extend this previous work to include new factors. Testing may be performed over a wider range of velocities and particle sizes than were previously available, allowing extension of models developed to predict damage. Currently, damage assessment in the high velocity regime must be extrapolated by use of such models. These could be improved and better calibrated by an EML based facility. The range of relevant research which could be accomplished by an EML based hypervelocity impact facility includes: 1) Simulation of hypervelocity impacts on components, 2) Evaluation of impact resistance of components, 3) Evaluation of improved bumper/wall shielding concepts, and 4) Penetration equation data for new shielding materials.

In addition, such a facility could also conduct basic research in hypervelocity impact phenomenon modeling, measure the equations of state of materials under very high pressures, examine methods of minimizing secondary ejecta and debris creation, and capture of orbital debris.

Currently, the experimental data base in these areas is limited, as research is

conducted using projectile velocities under 10km/sec. Preliminary investigations of these applications are underway at UT-Austin in conjunction with NASA/JSC. In order to develop a research facility, the railgun must provide projectile velocities in the range of interest. This is expected to occur as a result of current CEM research. Improved railgun diagnostic techniques would be required to allow for velocity measurement and observation of impact events. A combination of laser and x-ray based velocity sensors and an ultra-high speed camera system would be needed. This camera system would require 2×10^6 frames/sec, 5 nanosecond exposure time using a laser diode, 200-300 micron resolution (at 15 km/sec). Such a system would have to be custom designed and built, as it is at the state-of-the-art for hypervelocity impact research. Improved target diagnostic capability could be provided on-site and in user laboratories. These may include SEM, x-ray diffraction, and analytical tools (micro-probe).

An EML based research facility of the type described would have to be a multi-user facility in order to support the level of activity necessary. Preliminary efforts are underway to configure a railgun homopolar generator for impact studies.

Launchers and Accelerators

The railgun and coilgun can serve as the basis for a new family of launchers. When designed with a linear configuration, the coilgun can accelerate a mass with no direct sliding contact. This principle, combined with homopolar generator energy storage could allow launching of large masses (e.g. aircraft,

spacecraft). These would be useful for fully fueled vehicles where speed and weight are at a premium, such as short field take-off or vehicles that achieve orbit from runway launch. NASA is currently investigating the application of such launchers in advanced programs. Direct orbital insertion from the Earth or moon is a possibility.

Space Propulsion Applications

The same limitations which apply to chemically driven guns apply to chemical rockets. There are theoretical limitations imposed on all rocket engines based on thermodynamic principles. These include engines which burn propellants, and those which heat them (e.g. resistojets, nuclear thermal engines). Electric propulsion systems have been designed to overcome these limitations by using electric and magnetic fields to accelerate a working fluid.

These systems have found limited application due to their low thrust and low specific power. Applications have been further limited by the high mass power systems required. Application of the railgun-compulsator system to propulsion can overcome these limitations in certain applications. Relatively large amounts of working fluid can be accelerated as plasma to high velocity by the Lorentz forces, resulting in high specific impulses with higher thrust levels than previous electric engines. Specific impulses of 10,000 sec with thrust levels of 100 pounds are theoretically possible. A wide range of working fluids such as hydrogen, nitrogen, water, and inert gases may be used. Materials compatibility must be resolved, but experimental evidence suggests that these fluids will work. Smaller thrust levels (1-10 pounds) may also have

a variety of attitude control and station-keeping applications. The most likely application could be on spacecraft which require large (1MW) on board power supplies (weapons, active discrimination sensors) which can share components (radiators) with a propulsion system. This concept is illustrated in Figure 6-12, both in a schematic and in comparison with other propulsion systems. It can be seen that an EML based propulsion system has attractive weight and performance features.

The EML and PPT technology is brand new, as the compulsator was first developed at UT in 1978. Very little mission analysis has been performed which incorporates EML based propulsion. The availability of EML based thrusters could greatly enhance the ability of certain spacecraft (weapons platforms) to maneuver, thus enhancing survivability. For this application the systems must be evaluated for lifetime and reliability. If EML based propulsion is feasible and the opponent does not incorporate this technology, remaining in the chemical domain, a decided advantage may be obtained. The present state of Soviet development in this area is not known, but if evaluation and testing proceeds at a prompt pace, some advantage may be conferred to future American spacecraft with this capability.

Conclusion

This white paper has briefly discussed possible new applications of a still new technology. The authors suggest it may be in the interest of DOD to further capitalize on an investment in EML technology to meet a wide range of national security objectives.

References

The Acceleration of Macroparticles and a Hypervelocity Macroparticle Accelerator. Ph.D. thesis. Australian National University, Canberra, 1972.

Atkin, James, Analyses of EML Based Space Propulsion, Center for Electromechanics, Unpublished work, 1985.

Jones, A. H., W. M. Isbell and C. J. Maiden. "Measurement of the Very High Pressure Properties of Materials Using a Light Gas Gun." Technical Report TR65-84, November 1965, AC Electronics Defense Research Laboratories, Santa Barbara, CA.

Marshall, Richard A. "Earth to Space Launcher System." Publication PN-74, The University of Texas at Austin, Austin, TX, 1981.

Cour-Palais, Burton G. "Hypervelocity Impact Investigations and Meteoroid Shielding Experienced Related to Apollo and Skylab. NASA Conference Publication 2360, Orbital Debris. 1985.

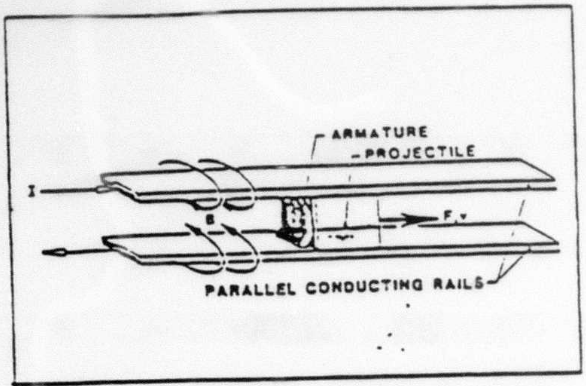


Figure 1. Electromagnetic Railgun

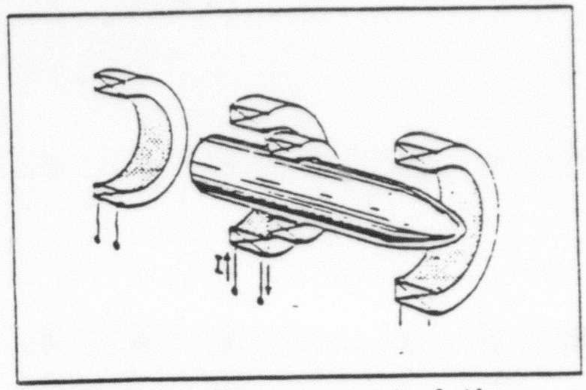


Figure 2. Electromagnetic Coilgun

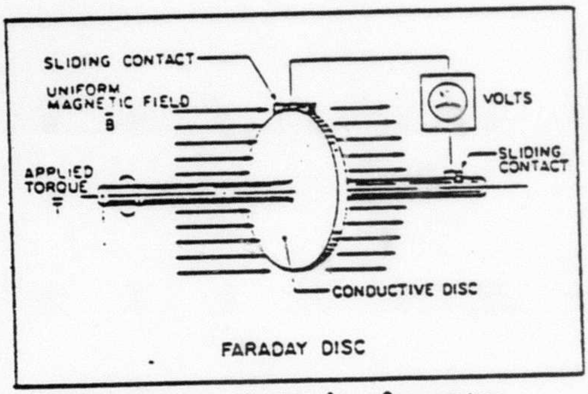


Figure 3a. Homopolar Generator

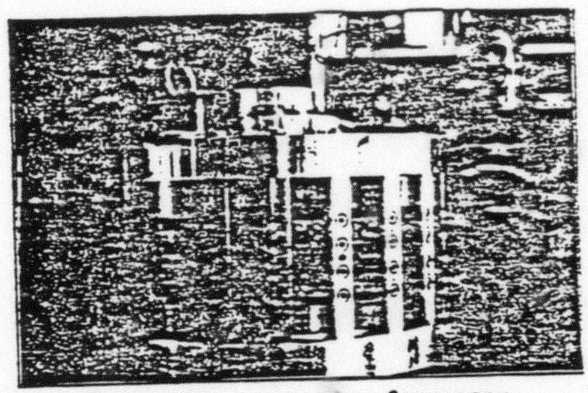


Figure 3b. Homopolar Generator

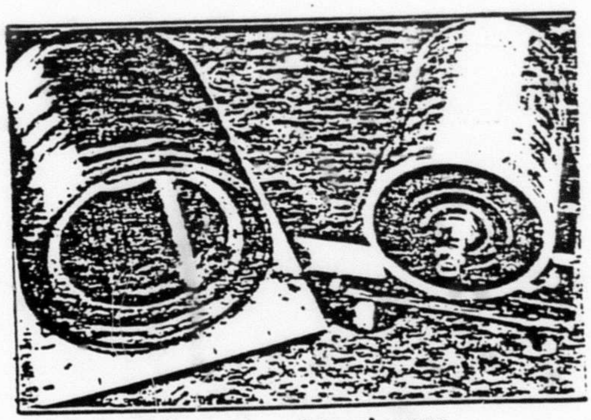
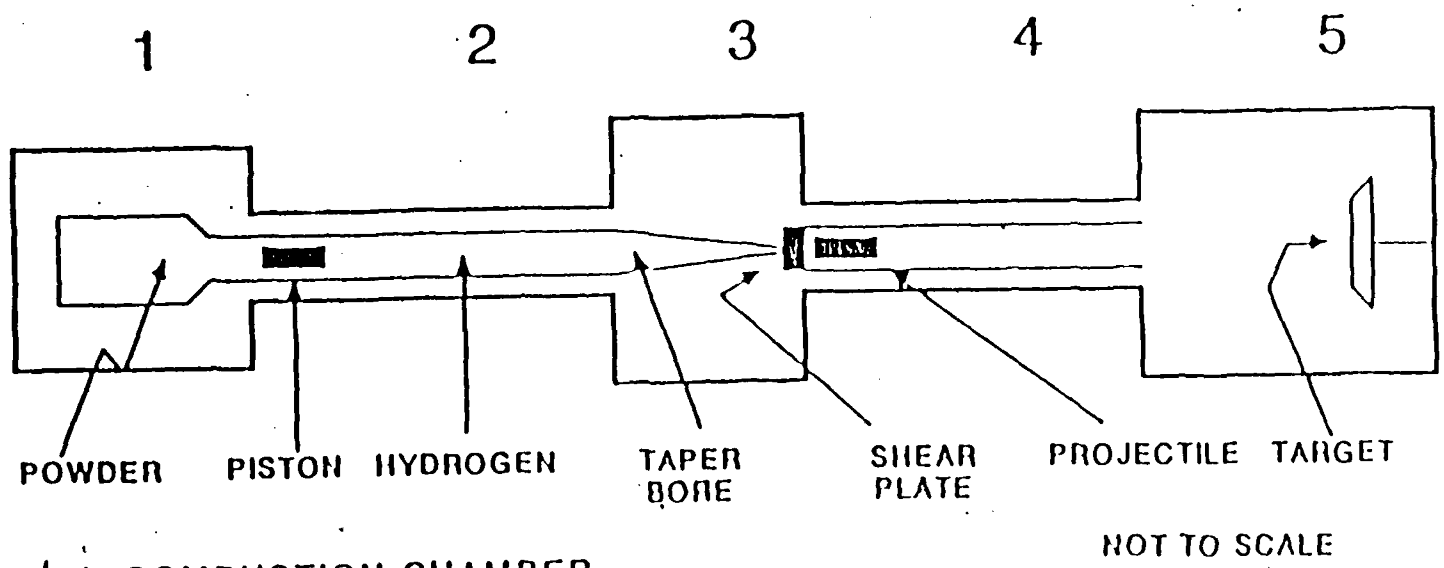


Figure 4. Compulsator

SMALL LIGHT-GAS GUN

GUN CONSISTS OF FIVE MAIN PARTS



1. COMBUSTION CHAMBER
2. PUMP TUBE
3. HIGH PRESSURE COUPLING
4. LAUNCH TUBE
5. TARGET CHAMBER

Fig. 5

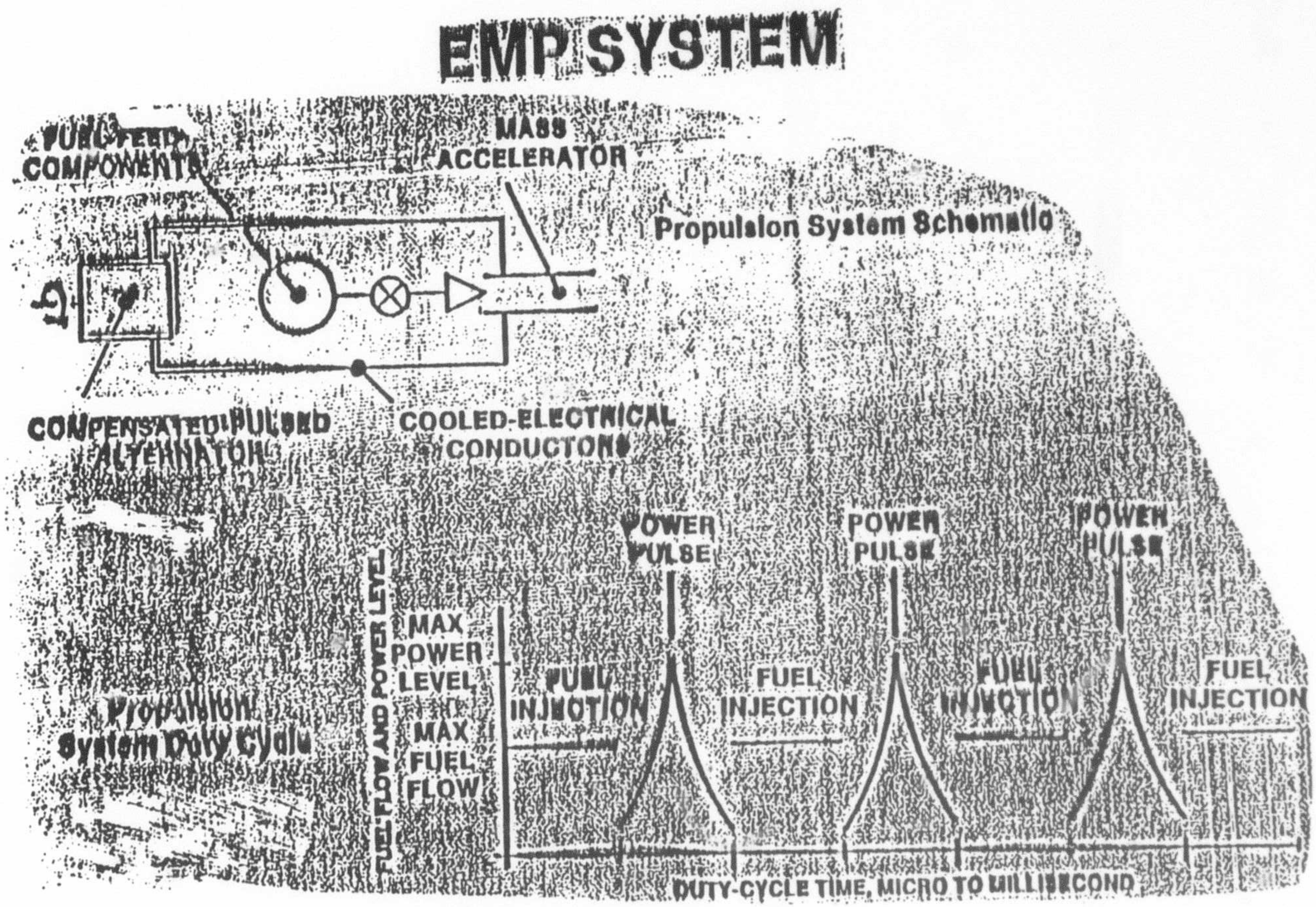


Figure 6

Figure 6 illustrates the EML thruster concept, showing how propellant and power pulses are fed to a thruster.

74

System Concepts Studied

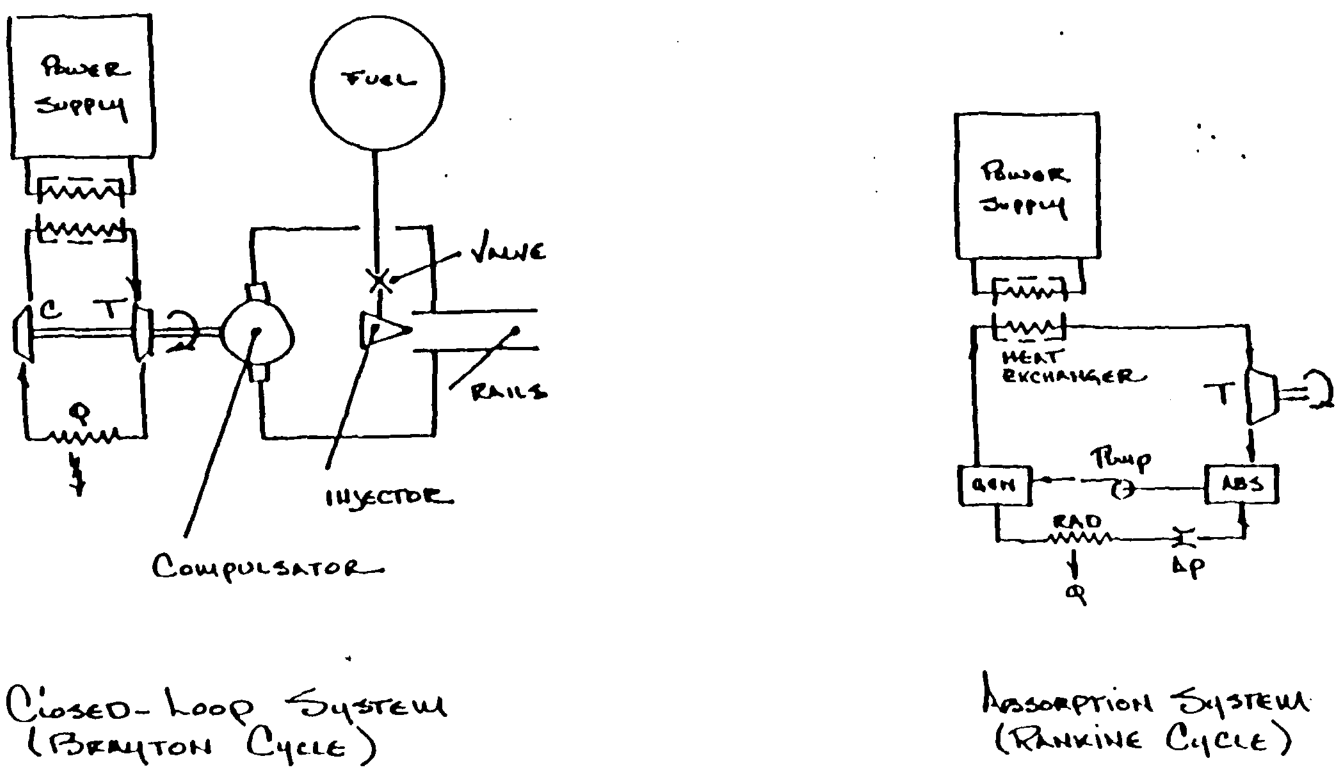


Figure 7

Figure 7 shows two alternatives for supplying pulse power to the EML system from a primary heat source on board the spacecraft. These are a closed loop Brayton cycle and an absorption based Rankine cycle. Some components (e.g., heat source and radiator) can be shared with other spacecraft systems.

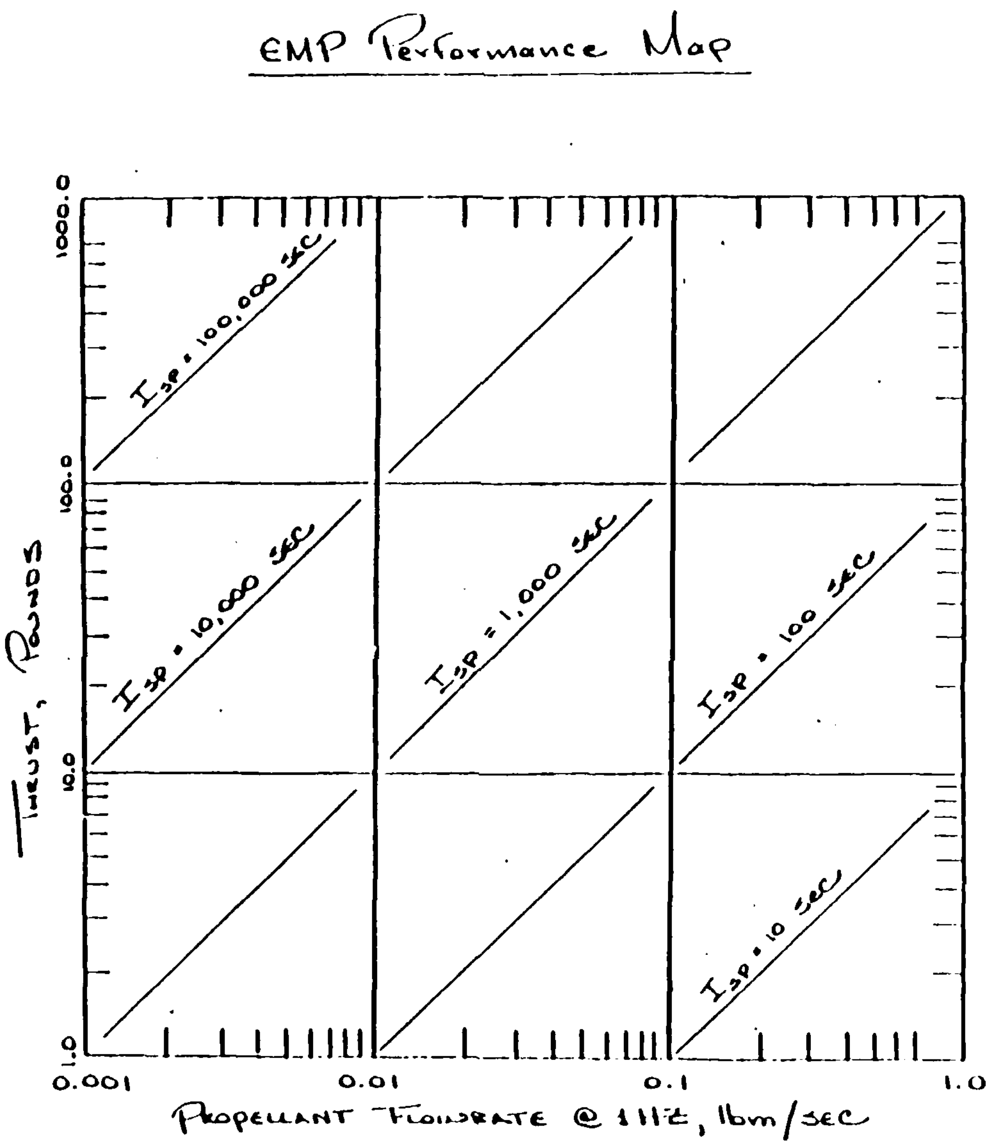


Figure 8

Figure 8 gives relations between propellant flow rate and specific impulse for a set of thrusts, illustrating the propellant requirements for an EML thruster.

EM PROPULSOR SIZE AND PERFORMANCE

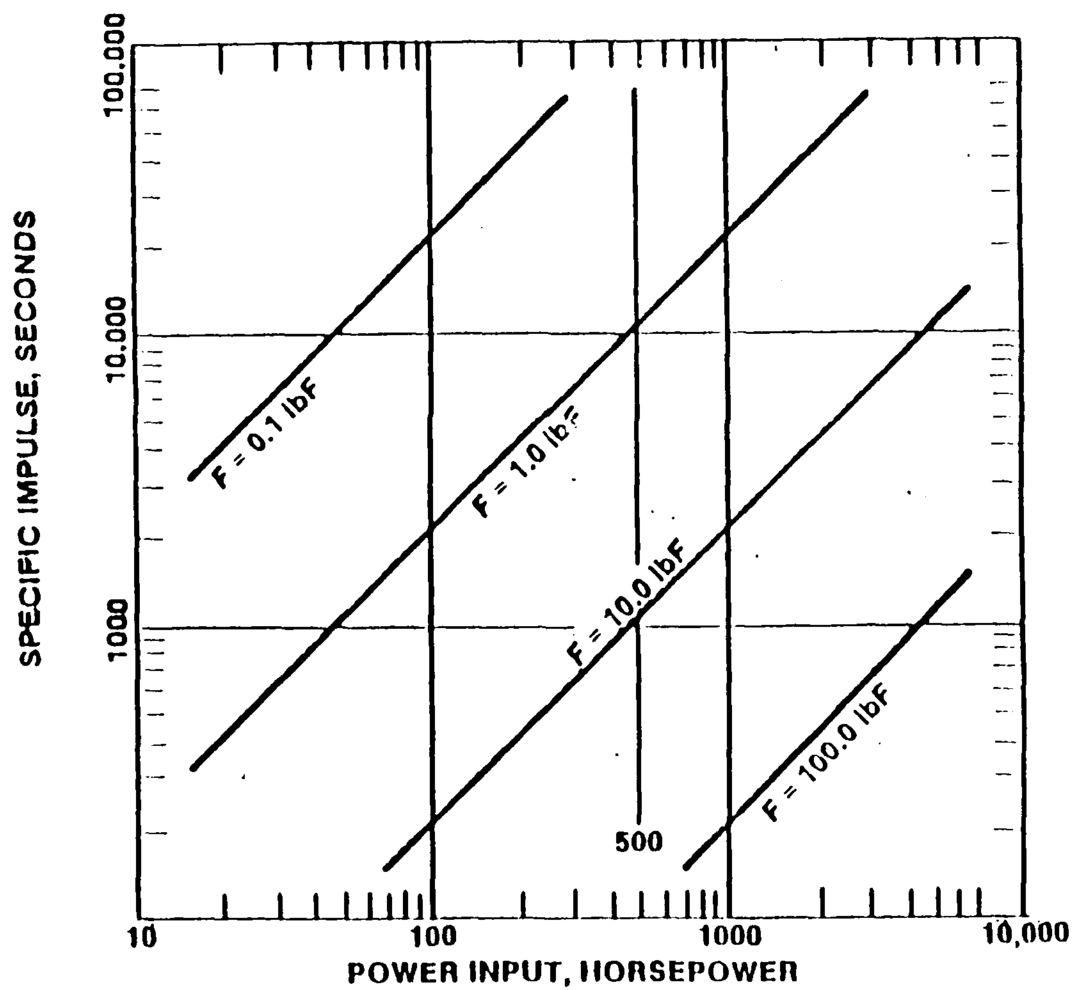


Figure 9

Figure 9 illustrates the power requirements for an EML based thruster in the thrust range from 0.1-100 lbF.

PROPULSION-SYSTEM WEIGHT ESTIMATE EML PROPULSOR(S) FOR ACS/STA. KEEP/DRAG MAKEUP FOR THE 300,000 LBM, 10-YR LIFE PLATFORM

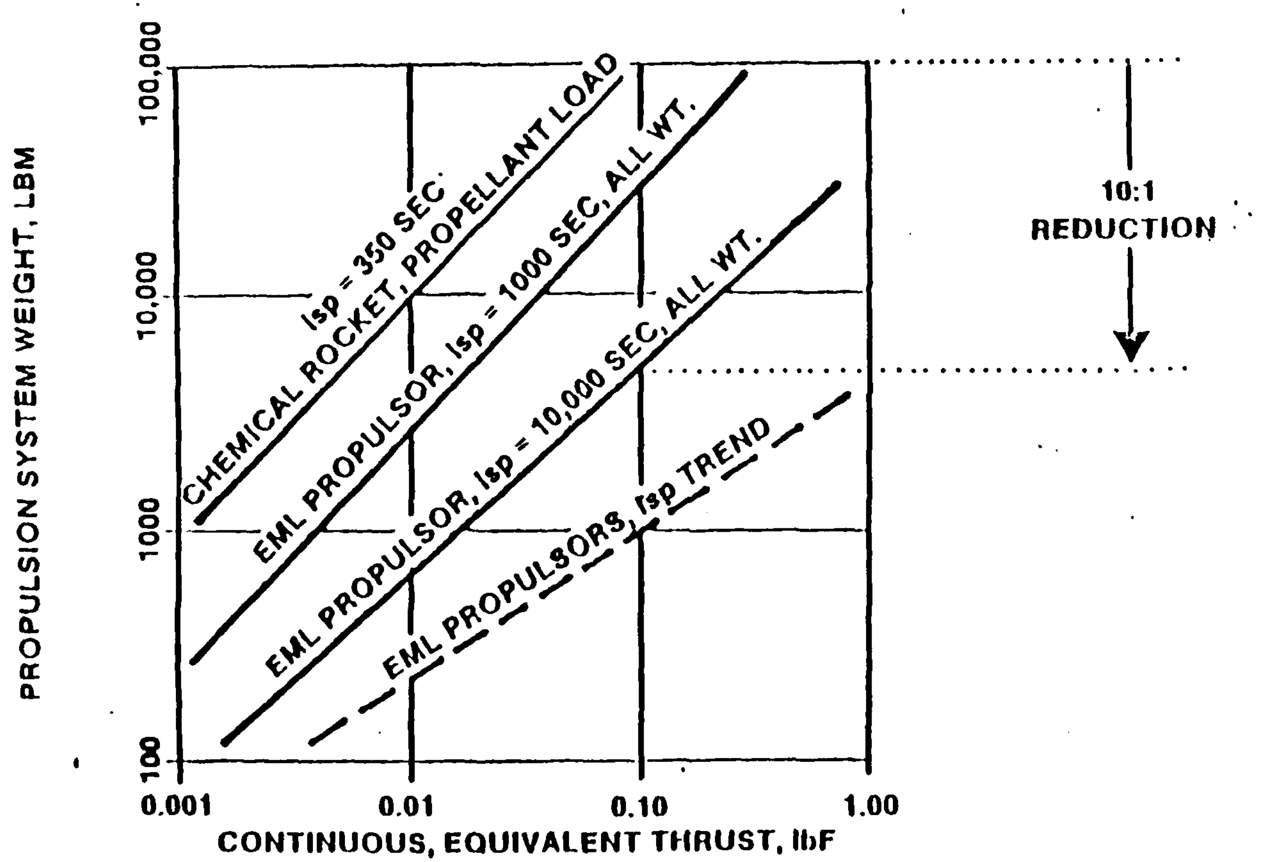


Figure 10

Figure 10 gives a weight comparison between chemical systems (including propellant) and EML systems which include all support mass (i.e., no mass shared with other systems) and the trend for mass sharing.

Propulsion-System Weight Estimate
EML Propulsion for the OTV Function

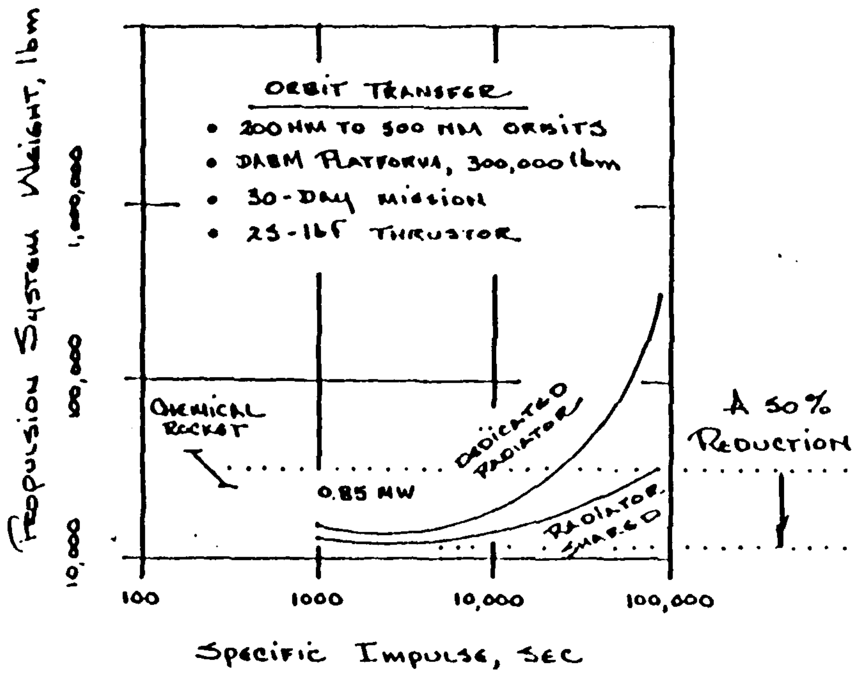


Figure 11

Figure 11 illustrates the mission application studied to date, a 300 Klbm space-based weapons system, showing the performance advantage of an EML system when components are shared and when they are not. In both cases a decided advantage is conferred to the EML systems.

ADVANCED PROPULSION CONCEPTS

Air Force Flight Test Center, Contractor Relations' Office, (BC), Stop 130, Edwards AFB, CA 93413 (JA Boucherle 805/277-2819)

PULSED PLASMA THRUSTER MISSION ENDURANCE TEST. A performance evaluation and life demonstration of the 4.45 millinewton thrust pulsed plasma propulsion demonstration and performance evaluation for the advanced AFRPL pulsed plasma propulsion system. A complete performance characteristics evaluation will be conducted seeking to generate important data validating the mission life capability of the pulsed plasma thruster design for a future Air Force Space Division Advanced Development Program. Overall system requirements will provide for a total impulse of 300,000 Newton-seconds at simulated space vacuum conditions. Ref purchase req FY1619-84-00138. Telephone requests will not be honored. See notes 17, 49, 68 and 88. (088)

A CDB Announcement
March 30, 1984

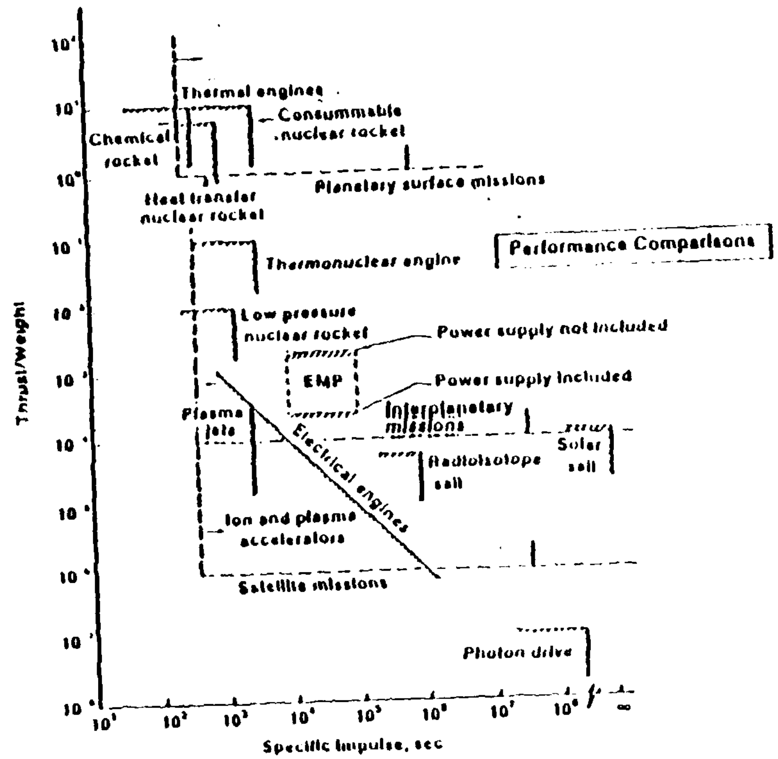


Figure 12

Figure 12 shows how an EML thruster system compares with other propulsion systems in both thrust/weight and specific impulse. An EML based system occupies a new niche in the parameter space.

LJI-R-85-348

STRATEGIC COMPUTING APPLICATIONS PROGRAM

BY

JAY BORIS*

AND

PETER HAMMERLING

NOVEMBER 1985

LA JOLLA INSTITUTE
P. O. BOX 1434
LA JOLLA, CA 92038
(619) 587-6000

THIS RESEARCH WAS SPONSORED BY THE
DEFENSE ADVANCED RESEARCH PROJECTS AGENCY
UNDER ARPA ORDER NO.: 3710
CONTRACT NO.: MDA903-85-C-0187

*Laboratory for Computational Physics, Naval Research Laboratory

TABLE OF CONTENTS

I. EXECUTIVE SUMMARY

The following paragraphs describe a Strategic Computing Applications Program (SCAP) to be considered for support by DARPA. The efforts of the program are distributed in a cross-disciplinary way over several program areas in DARPA and make specific military use of the hardware and software developments in DARPA's Strategic Computing Program. The operational goal of the program is the orderly, timely, and effective use of the parallel computer architectures being developed under several DARPA projects in, and related to, strategic computing. Through this use a spectrum of important, defense-related problems will be given early access to promising advances in computing and the software and hardware developers will get valuable user feedback. Substantial immediate value will be derived from the initial development, testing, and application of these advanced systems long before benefits from any intended end use accrue.

The focal project of the program is the development and testing of efficient, reliable, transportable, processing elements to be carried on-board a platform (satellite, ship, aircraft, or land vehicle) that could perform Surveillance, Correlation, and Tracking (SCAT) computations on multiple targets of interest and that would interface these new capabilities to human decision makers. This technology insertion program requires the highly parallel computer architectures already being developed since surveillance, correlation, and tracking is a computationally intensive problem. The Monotonic Logical Grid (MLG) algorithms, described below and in Appendix B, promise up to two orders of

magnitude speedup in major, compute-intensive aspects of the SCAT problem. Combined with a factor ten speedup using the new strategic computer architectures--the called for three orders of magnitude speed improvement is here now.

Initially, of course, a "breadboard" simulation of the proposed devices would need to be demonstrated. Another application would be as a simulator for battle management strategy and policy development. The latter could be developed in the context of a user facility and the experience gained in its implementation and use could be transferred to specific service-oriented missions embodied in the modules mentioned above.

The SCAP proposed has three components, a concerted attack on the omnipresent SCAT problem and two closely allied advanced research programs: one on advanced computer architectures and the other on relevant scientific applications.

All three components of the program will be hosted on the NRL Cray XMP, the TMC Connection Machine, and the other facilities of the NRL Central Computer Facility, and the Laboratory for Computational Physics. This program would involve all components of the Heterogeneous Element Supercomputer System (HESS) being assembled at NRL and will incorporate DARPA R&D efforts which can contribute to the SCAT problem or to the other aspects of the Strategic Computing Program described above. In this regard, the HESS would function and be maintained for DARPA/DoD users much as the Numerical Aerodynamic Simulator is being developed for NASA.

The cost of the overall SCAP will be about $40M for five years based on incorporating ongoing (funded) work which would benefit by being coordinated and some new work in areas of high possible return which depends synergistically on the unique HESS capability. The program should occur in three phases:

I) Planning Phase - 1 year ($2M)

II) Development Phase - 3 years ($7M, $9M, $12M)

III) Production Phase - 1 year and follow-on ($10M/year ...)

This third phase may involve substantial new equipment both for second generation Strategic Computing systems that extend the capabilities of HESS as well as to provide useable components for field trials and extended breadboard experiments. The program would provide some software development support to enhance the usefulness of the new hardware components.

This Strategic Computing Applications Program deserves a prominent place in the DARPA "Strategic Computing" (and the SDIO programs) for four important reasons:

1) There has been an algorithmic breakthrough, a data structure called the Monotonic Logical Grid (MLG) which allows efficient parallel processing of manybody interaction problems. This innovation deserves extensive development and optimization in the context of Battle Area Management, the SDI missile defense management problem, physical manybody problems, cellular automata, fluid dynamics, line-of-sight graphic displays, and symbolic data organization. The MLG technique for organizing constantly changing data dynamically is

85

capable of optimal execution on the SCAT problem using parallel architectures such as the TMC Connection Machine, currently being developed for DARPA by Thinking Machines Corporation, the Cray XMP and the GAPS, a vector of array processors in place already at NRL.

2) The community of researchers and users which this program would draw together around the HESS would provide necessary practical experience bridging current gaps between signal processing, artificial intelligence, and numerical simulation. All three components are necessary in an autonomous vehicle or device, for example. Signal and image processing are required for the digestion of sensor information about the environment of an autonomous vehicle or system. This information would be correlated and evaluated by the A.I. central inference engines and associative memory pattern recognizers. Numerical simulation is required to estimate environmental variations and to project the likely outcome of possible decisions for use in the computerized decision-making processes.

3) Existing DARPA and SDIO programs and projects need substantial raw compute power which the HESS can provide cost effectively for important classes of military problems. The NRL Central Computer Facility is also available to service off-site DARPA contractors. The Laboratory for Computational Physics and the Naval

Research Laboratory will provide program coordination and support as well as fertile ground for colleagial interactions and technology transfer.

4) The existence of VOYEUR, an interactive, high-bandwidth, color graphics system at NRL, allows immediate analysis and understanding of simulations and calculations being performed anywhere on the HESS as they are being performed.

These four reasons address important requirements of the DARPA Strategic Computing Iniative and the SDIO program. Support is requested for the program as a whole to emphasize and encourage the interaction of the distinct parts. A technical "board of directors" for the program would be devised consisting of technical experts in the three areas and representatives of the major sponsoring components.

II. BACKGROUND

As a community we do not know a lot about writing concurrent programs. There has been progress in ten years, primarily in the physical and engineering sciences, but we still have lot to learn before massively parallel processing is easily and commonly used in production environments. This kind of software seems to be an order of magnitude more difficult to build and maintain than conventional software. Algorithms to use concurrent, distributed-processor architectures will generally take as long to perfect as the hardware, typically several three-five year computer generations.

For several important classes of problems, however, we have been able to develop multiply parallel algorithms which lend themselves to particularly efficient implementation on a distributed, high-concurrent processing system. These problems include basic investigations into cellular automata and nonlinear system dynamics, simulation of large physical systems such as fluids, reacting flows, molecular dynamics, etc., direct scattering calculations such as occur in acoustics and radar, and the militarily pivotal SCAT problem. A number of existing DARPA and SDI-related projects also require large amounts of time on a supercomputer with millions of words of fast memory. Following is a partial list of projects which are currently underway and which require or could use the HESS profitably.

- Battle engagement area Surveillance, Correlation, and Tracking (SCAT) for thousands of targets, whether air, land, sea, or space, requires small stand-alone systems

with supercomputer capability. Significant portions of the following text address these generic issues in the specific context of the SCAT problem. The newly-developed, parallel MLG algorithms provide an efficient dynamic basis for organizing and retrieving SCAT data in real time.

- The advanced battle management systems will have to incorporate integrated aspects of all three types of processing, A.I., signal, and image processing, and highly-parallel numerical simulation techniques.

- Charged Particle Beam (CPB) secondary electron propagation requires more memory than available on a Cray 1 and the equivalent of several hundred Cray hours per year. The software technology to write a multidimensional (4 or 5 dimensions) Boltzmann code exists but has not been worth implementing because of the paucity of sufficient computation time.

- CPB hole-boring and anomalous channel-cooling simulations require multidimensional reactive flow production calculations on a Cray class system. The TMC Connection Machine would be particularly effective in concurrent use with the Cray to more accurately track interfaces between materials and phases in high power-density systems.

- Advanced engine development using ramjets and scram jets requires long, expensive simulations in which the

acoustic waves, as well as subsonic and supersonic flows, are all resolved. The Cray XMP is adequate for the fluid flow calculations but the Connection Machine with a useable floating point capability is needed for detailed multidimensional runs with a real chemistry.

- Associative Memory (AM) research requires the massive parallelism which the HESS will provide through the Connection Machine and perhaps other "neural net" type architectures. Attention focusing, problem encoding, and fault tolerance are all better studied in large, general-purpose computer networks with interactive, high-bandwidth graphics systems. $N \ln N$ encoding and reading algorithms for AM systems must be found.

- Massively parallel processing also makes direct scattering computations, both acoustic and electromagnetic, possible for complex shapes and tailored surface material properties. The exact response of a given shape or material to new electromagnetic pulses can be calculated directly using a formalism developed by West, et al.

- Advanced A.I. systems for autonomous vehicles will require highly parallel distributed computation for prediction, signal, and image processing. Cooperating expert systems can be implemented within the HESS Network and have several high-powered compute engines available for those aspects of composite real problems which have a severe computational load.

- Damage and Lethality Research, whether from beams or kinetic energy weapons, requires molecular dynamics calculations on thousands of interacting particles in the "nearest neighbors" approximation. The MLG algorithm was originally designed for this purpose. It is ideally suited for massively parallel processing such as can be conducted with the Connection Machine.

- Biomolecular engineering promises a whole new family of special-purpose/high-performance materials of great military potential. DARPA is supporting much research in this field. The Monotonic Logical Grid (MLG) methods make possible truly large simulations with millions of particles to bridge the gap between microscopic atomic particles and macroscopic materials properties.

- Propagation of EM pulses over long distances through turbulent and disturbed media requires enormous parallel computations. New systolic architectures should be nearly ideal for these types of problems. Perhaps fed asynchronously by the Cray calculating the specific realizations of the disturbed environment through which the pulses are passing, the Connection Machine could be used at full effectiveness advancing nodes on a phase front passing through the disturbed medium.

- The same kind of synergistic interaction is needed for X-Ray Laser modeling and analysis. The Cray Class

Supercomputer is quite adequate for the fluid and thermal conduction dynamics which occur in physical scenarios but the complex atomic-radiation chemistry occurring at each grid point requires the massive parallelism of a Connection Machine. The separate processors could work together, each solving the part of the problem for which they are the most suited and exchanging only the information needed to continue the asynchronous but coupled computation.

- The potential of computer systems which are commercially available to our adversaries should be assessed for special and classified applications. This includes in particular the new bussed multi-array processor systems which require considerably more programming than a CRAY or VAX but which can also be far more cost effective. The HESS, through the Laboratory for Computational Physics' Graphical and Array Processing System (GAPS) also has a strong capability of this third kind, a hierarchical concurrent distributed processing system.

III. APPROACH

The Stratic Computing Applications Program, as planned, is composed of three parts, a focal military problem and two related advanced research areas. The focal problem is the SCAT problem and the related technology research areas are Advanced Computer Architectures and Relevant Scientific Applications.

The SCAP takes advantage of the HESS being assembled at the U.S. Naval Research Laboratory. The actors will be the members of the Laboratory for Computational Physics (LCP) and the Naval Research Laboratory, DARPA's applications contractors, and the strategic computing research communities. We expect some indirect benefits from carrying out this cross-disciplinary program. Several different scientific and technology communities will be brought into contact and held there long enough for the necessary interactions to explore the potential and exploit the power of the newly-developed computer capabilities.

The SCAP as proposed would provide for the integration of several rapidly evolving information processing technologies to tackle and solve one of the crucial issues of conflict and battlefield management, information logistics. The approach will be to provide at least one facility where users can develop and use software systems in conjunction with newly developed architectures. The HESS at NRL would be the primary system and would integrate the DARPA-developed Connection Machine of TMC, the Cray X-MP/12 at NRL's Central Computer Facility, the Graphical and Array Processor system (GAPS) in the LCP Physics and other special

systems as the users or DARPA might see fit to provide for use or comparison.

The three major areas of the Strategic Computing Applications Program are illustrated in the following diagram:

Computing Architectures	Surveillance, Correlation, & Tracking	Scientific Applications
Associative Memory Neural Networks	CRAY XMP & TMC Connection Machine	Physics & Mechanics Applications
Signal & Image Processing	Artificial Intelligence	Cellular Automata & Nonlinear Dynamics

At the center of the plan is the heterogeneous element supercomputer system being developed at the NRL Central Computer Facility.

The following brief outline, and those in Section V below, list a number of possible, high-leverage tasks which ought to be considered for support under each of these three program components. It would be the immediate task in the first year of the program to select a mix of problems and performers and to establish a program review and guidance board from among the technical performers and major sponsors, and to begin to implement the project.

Surveillance, Correlation, and Tracking Problems:

1. Three- and Four-Dimensional MLG "Tanks," Construct and Test
2. Cray - CM Integration, Software for Asynchronous Heterogeneous Systems
3. Develop Network and Simulation Programs, User and Engineering Support
4. Make CM Available to DARPA Users, User and Engineering Support
5. Cooperating Expert Systems, etc. Algorithms for SCAT Test Bed
6. Fault Tolerant Distributed Data Base Techniques
7. Develop Interfaces for Hardware Systems, etc., NRL Test Bed Program

One approach to solving a new and complex problem is by large jumps, major deviations from accepted practice with the promise of big dividends and with correspondingly large risks. It is often the case that similarly big dividends can also be had by a planned series of steps building on known principles and accepted practice. Such is the approach we are proposing here. Signal processing, computational physics simulations, and A.I. are growing together in several important ways and this project takes heed and advantage of these confluent trends.

As discussed below, a three-order-of-magnitude quantum jump will be provided by implementation of a unique software development, the Monotonic Logical Grid on the HESS. Coupled with

new computing architecture (e.g., the CM) one can design and construct a large-scale simulator.

Strategic Computing Applications Program to Demonstrate a 3D/4D Battle Engagement Area Simulation Tank

The SCAT project will first implement a prototype Battle Engagement Area Simulation Tank (BEAST) in three dimensions as we have discussed. This would prepare the way for the 4D tank needed to sort the contact reports for geolocality before fusion, correlation, and identification analysis, etc. This tank would be used by the "Dungeon Master" and/or a battle commander to visualize the current configuration as a dynamically evolving data base. It would be a necessary component of an overall simulator and a test bed for the 4D BEAST which would follow on the Connection Machine in Fy 86 or 87. It would also provide, almost immediately, a driver for developing the systems and software to display and interact with the evolving scenario.

The Monotonic Logical Grid

Central to our concept is the use of a new dynamic data structure called the MLG which was devised specifically to organize the geometric information describing randomly moving nodes for parallel computation. The technique arranges the data for the nodes in computer memory in a simple mapping which ensures that near neighbors in space are automatically near neighbors in the data arrays as well. This technique is central to the proposed attack on the SCAT problem because it forms the basis of the 3D scenario tank and the 4D correlation tank used to identify, weed out, and collapse the contact reports in real time. The

MLG-based BEAST task is addressed and references are given in the Appendix A.

The MLG reduces the number of interactions which must be considered for each new contact report from hundreds or thousands (nominally each evolving track) to a few (or a few dozen) by automatically grouping tracks and contact reports which are adjacent in space-time close together in the data structures. When the data giving the location and properties of the nodes is shifted around through the computer memory arrays of the MLG in a discontinous way which, however, maps monotonically onto the real motion. Substantial experience has been obtained in using the MLG to solve manybody problems with thousands of randomly-moving nodes and much more stringent accuracy constraints than required for SCAT.

By using the MLG, the contact reports will automatically be organized for parallel computation of report-report separations and track-report correlations. The combinatorial problem is reduced from an order N^2 problem to an order N problem. In practical terms this is a factor of 20-50 reduction in the amount of computing to be done for 1000 nodes and factors of 100-250 reduction when $\sim 10^4$ nodes are involved. Furthermore, the MLG happens to organize the computational problem data structure in a way ideally suited to the TMC "Connection Machine" being built for DARPA. This confluence of algorithmic and hardware capabilities should allow $\sim 10^6$ contact and track reports to be processed simultaneously with new reports coming in at the rate of 3000-5000/second. New techniques in AI and AM for track

identification and correlation will be employed under less stringent circumstances than otherwise would prevail because a large component of the identification and correlation problem, identifying only the near neighbor reports as candidates, is automatically accomplished by the MLG. Further details may be found in Appendix B.

The following figures, together with those in Appendix A, summarize the above concept.

WHERE IS THE "QUANTUM JUMP"?

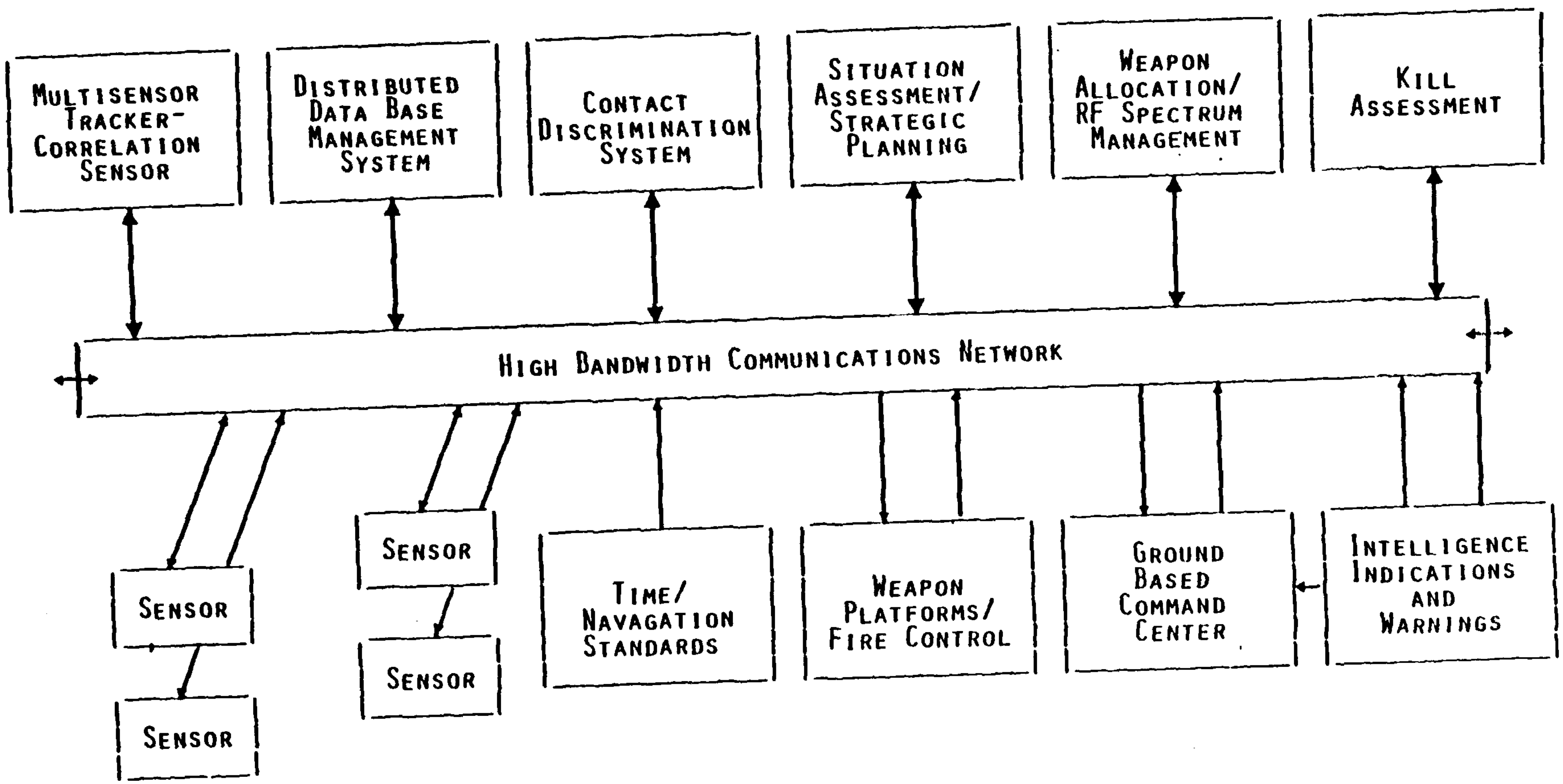
1. THE MONOTONIC LOGICAL GRID

A NEW ALGORITHM = FACTOR OF 100

2. ADVANCED STRATEGIC COMPUTING ARCHITECTURES

MASSIVELY PARALLEL HARDWARE = FACTOR OF 10

BATTLE MANAGEMENT TESTBED



4D "TANK" -- $[x_i, y_i, z_i, t_i]$

- MULTISENSOR TRACKER - CORRELATOR
- CONTACT DISCRIMINATION/SENSOR FUSION
- ORGANIZED, ANALYZED TRACK HISTORIES
- DISTRIBUTED DATA BASE MANAGEMENT KERNEL

3D "TANK" -- $[x(t), y(t), z(t)]$

- WEAPON ALLOCATION/RF SPECTRUM MANAGEMENT
- SITUATION ASSESSMENT/STRATEGIC PLANNING
- WEAPON PLATFORM FIRE CONTROL/KILL ASSESSMENT
- GROUND BASED COMMAND CENTER/DM

BATTLE ENGAGEMENT AREA SURVEILLANCE TANK

AN ACTIVE, RANDOM ACCESS DATA STRUCTURE WHICH CONTINUALLY ADAPTS TO THE CHANGING CONFIGURATION OF NEW CONTACT REPORTS, KNOWN COMBATANT MOTIONS, AND EVOLVING SENSOR TRACKS. GEOPROXIMITY OF NEIGHBORING DATA NODES IS ASSURED BY OPTIMUM PARALLEL PROCESSING TECHNIQUES.

Generic Problem--BEAST Goals

$\sim 10^5$ objects at once	64K 3D MLG Tank
$\sim 10^7$ contact reports at once	1024K 4D MLG Tank
$\sim 10^4$-10^5 reports every 10 sec	"digest" 10^3 reports/sec
$\sim 10^3$ global updates/hour	5-10 updates/min

Cost of Nearest-Neighbor Algorithms

Direct Vector Summation:

$$\#\text{OPERATIONS} = 7.5 \times 10^8 \rightarrow \sim 15 \text{ SECONDS/STEP}$$

$$= 5000 \text{ PTS} \times \frac{5000 \text{ PTS}}{2} \times \frac{60 \text{ OPERATIONS}}{\text{PT}\cdot\text{PT INTERACTION}}$$

Monotonic Logical Grid Algorithm:

$$\#\text{OPERATIONS} = 2.16 \times 10^7 \rightarrow \sim 0.43 \text{ SECOND/STEP}$$

$$= 5000 \text{ PTS} \times 60 \text{ NEAR PTS} \times \frac{60 \text{ OPERATIONS}}{\text{PT}\cdot\text{PT INTERACTION}} \qquad 3600$$

$$+ 3 \text{ DIRECTIONS} \times \frac{4 \text{ ITERATIONS}}{\text{DIRECTION}} \times \frac{60 \text{ OPS/PT}}{\text{ITERATION}} \qquad 720$$

/0-

IV. THE STRATEGIC COMPUTING APPLICATIONS PROGRAM FOCAL PROBLEM

The proposed technology insertion program has, as its focus, military SCAT since this computationally intensive problem occurs essentially universally in battle management wherever numerous mobile units and platforms are involved. A number of different important applications exist--maritime surveillance from satellites, the outer air battle, surface and subsurface ocean forces, fleet defense, SDI missile scenarios, kinetic energy weapon targeting, land combat simulation, air traffic control, etc.--which have in common the following major difficult aspects:

1) many separate objects moving rapidly in complicated paths (tracks),
2) numerous observations (contact reports) of the same object by different sensors at different times and with different errors,
3) need for fault tolerant input, performance, and analysis components, and
4) large, asynchronous, data flow and dynamic, distributed data base requirements.

To deal with these aspects rapidly and efficiently calls for both the appropriate hardware and software, such as discrimination and correlation algorithms, as well as a computationally efficient method for logically following all targets of interest.

There are seven stages to the SCAT problem:

1) Beam Forming--The raw data input to a given sensor is developed into a current contact report list for the sensor.
2) Contact Report Acquisition--Contact reports of different character, frequency, accuracy must be assembled, verified, and weeded.
3) Multisensor Integration--A preliminary sort of new contact reports is prepared for insertion into the 4D MLG tank.
4) Correlation Data Collapse--Multiple images are identified and collapsed down to identifiably different contact reports.
5) Track Identification--The system must extend previously identified tracks into new data and develop new tracks.
6) Platform Vector Updates--The system must satisfy information requests on vicinity of a given track or platform in burst mode.
7) Ordered, Analyzed Track Histories--These have to be stored dynamically so they can be retrieved and updated simply.

Not all seven states make direct use of the Monotonic Logical Grid "tank" data structures. Beam Forming and Contact Report Acquisition prepare data in a form suited to using the tanks. Multisensor Integration and Correlation Data Collapse are built around the MLG algorithms as is the seventh stage, Ordered, Analyzed Track Histories. Track Identification and Platform

Vector Updates are not performed in or by the tanks but gain appreciably from the format of the adjacent reports and tracks used by the MLG.

V. COMPUTING ARCHITECTURES AND SCIENTIFIC APPLICATIONS

Two important auxiliary themes are included in the overall SCAP to provide a spectrum of interested, available expertise to tackle the focal problem and to provide additional channels for productive use of the new architectures. These two auxiliary themes are Relevant Scientific Applications and Advanced Computing Architectures. Each has a limited number of focused efforts on specific topics which also contribute to the focal theme of providing a Battle Engagement Area Surveillance Tank with hooks where different components of the SCAT problem can be introduced and tested in an adaptable, well-diagnosed environment.

Advanced Computing Architectures:

1) N ℓn N coupled AM
2) Cellular Automata Models of Physical Operating Systems (CAMPOS)
3) Dynamic Data Structures
4) User Software (BASIC or C)
5) HESS
6) Pattern and Speech Recognition

Relevant Scientific Applications:

1) Direct Scattering
2) Microwave Pulses
3) Multimaterial Interface Tracking
4) Chemistry Integrators for Independent Stiff Systems
5) Computational Fluid Dynamics
6) Biomolecular Engineering

VI. RESOURCES

As outlined below, the program would provide for the integration of several rapidly-evolving, information-processing technologies to tackle and solve one of the crucial issues of conflict and battlefield management: information logistics. The approach will be to provide at least one facility where users can develop and use software systems in conjunction with newly-developed architectures. The HESS at NRL, which is available to serve offsite DARPA contractors, would be the primary computational system and would integrate the DARPA-developed Connection Machine of TMC, the Cray X-MP/12 at NRL's Central Computer Facility, the GAPS in the LCP, and such other special systems as the users or DARPA might see fit to provide. As prototypes of other advanced machines become available, e.g., one based on neural network models, they could be interfaced with the rest of the hardware and made available on the NRL broadband network.

The cost of the overall program will be about $40M for five years and will be incurred in three phases:

I) Planning Phase - 1 year ($2M)

II) Development Phase - 3 years ($7M, $9M, $12M)

III) Production Phase - 1 year and follow-on ($10M/year ...)

This third phase involves new equipment for second generation super computer systems to extend the capabilities of the HESS.

VII. SUMMARY

The Surveillance Correlation and Tracking Problem is central to the management and use of a number of military systems. Each version of the problem has different complicating aspects and computational bottlenecks but the problem similarities outweigh the differences. World-wide ocean surveilance, theatre-wide maneuver coordination, the outer air battle, and the SDI endoatmospheric identification and targeting scenarios all require complicated analyses of a very large number of contact reports to sort out a large number of real sources, the false sources, and redundant observations of the same source by different sensors.

A usable, multipurpose breadboard simulation "tank" for battle engagement area modeling can be developed using resources currently or soon to be available on the NRL network. Ultimately, specialized modules can be developed for use on various platforms (satellite, air, sea, or land) to provide autonomous, on-station performance and/or visual, real-time information and support to human decision makers.

The full SCAP deserves a prominent place in the DARPA "Strategic Computing" program for four important reasons.

1) The MLG technique for dynamically organizing constantly changing data is capable of optional execution on such a system and requires extensive development and testing in the context of battle area management, physical manybody problems, line-of-sight problems, and symbolic data organization.

2) The HESS would provide experience bridging some of the existing gaps between signal processing, artificial intelligence, and simulation. All components are necessary in an autonomous vehicle or device.

3) Existing programs and projects need substantial raw compute power which such a system can provide quite cost effectively for a limited but important class of problems.

4) New methods of interactive display of evolving data in multiply parallel systems must be developed. VOYEUR is a major step forward.

These four reasons address important aspects of the SDI program and the Strategic Computing Iniative, thus support is requested for the program as a whole to emphasize the interrelationship of the distinct parts.

APPENDIX A: A MONOTONIC LOGICAL GRID SIMULATOR - THE BATTLE ENGAGEMENT AREA SURVEILLANCE TANK

by Jay P. Boris and J. Michael Picone, NRL, Code 4040

The Surveillance, Correlation, and Tracking Problem (SCAP) is our focal theme since this computationally intensive problem occurs essentially universally in military engagement planning and battle management wherever diverse mobile units and platforms are involved. A number of different important applications exist--world-wide maritime surveillance from satellites, theatre-wide maneuver coordination, the outer air battle, surface and sub-surface ocean forces, fleet defense, the SDI endoatmospheric identification and targeting, kinetic energy weapon scenarios, land combat simulation, air traffic control, etc. Each version of the problem has different complicating aspects and computational bottlenecks but the problem similarities outweigh the differences. These problems have in common the following major difficult aspects:

1) many separate objects moving rapidly in complicated paths (tracks),
2) numerous observations (contact reports) of the same object by different sensors at different times and with different errors,
3) need for fault tolerant input, performance, and analysis components, and
4) large, asynchronous, data flow and dynamic data base requirements.

A usable, multipurpose breadboard simulation "tank" for battle engagement area modeling can be developed using resources currently or soon to be available as part of the NRL Central Computer Facility (CCF), a network-based Heterogeneous Element Supercomputer System, an the Laboratory for Computational Physics' new Graphical Array Processing System (GAPS). The major component of a Surveillance, Correlation, and Tracking (SCAT) system based on the Monotonic Logical Grid (MLG) algorithms are shown in Figure 1. The use of the MLG to organize the inflowing contact reports and sort out the large number of real sources from the many false sources and the redundant observations of the same source by different sensors, is described below. This Battle Engagement Area Surveillance Tank (BEAST) project, based on multidimensional MLG data structures, serves several important military purposes:

1) It would develop a 3D MLG monitoring "tank" in which the location and relationships between evolving tracks and identified participants can be maintained and updated in real time. This capability serves to define the engagement area and would be a central software component in all engagement area simulations and real-time correlation analyzers.

2) It would develop a 4D MLG sorting "tank" for sorting and analyzing the relationships between incoming contact reports, evolving tracks, and identified participants in real time. This tank would contract the data base

dynamically as tentative correlations and track identifications are made. This feature distinguishes the proposed simulator from any other battle management system.

3) It would provide a variable-parameter, monitored engagement area simulation capability for testing algorithms, software systems, and even new hardware interfaces. This capability would also be pushing back the envelope on performing the necessary computations fast as needed as parallel processing breakthroughs can be accommodated.

4) It would provide a valuable data base on distributed, fault-tolerant system integration to accomplish complex, military objectives.

5) It provides a simple, systematic method for archiving the incoming data and the resulting identifications/ decisions. This capability allows rapid backtracking when valuable, time-late data are received and follows from the structure and dynamics of the 4D MLG sorting and analysis tank.

There are several levels of simulation involved in this project. Simulations of engagements can be run on the entire system with the GAPS 3D tank developing the opponent scenario and monitoring the analysis and decision-making procedure. The system could be used to perform confrontation and battle simulations as a training device. It could also be used to simulate an operational

system or to test peripheral component systems such as sensor interfaces, distributed data base updates, etc. in a relatively realistic environment. The BEAST could also be used to develop strategy and policy regarding possible future systems. The BEAST will be a breadboard for the computationally intensive parts of a real battle management system, an engineering simulation of the system that would be used in a real engagement. In particular, the speed of critical algorithms could be assessed to supercomputer standards and in parallel computing environments.

The new MLG approach, developed by the Laboratory for Computational Physics for molecular dynamics, provides a new tool for attacking this class of problems. It simplifies the analysis of the overall SCAT problem because it provides a consistent viewpoint on the necessary components of the system and how they must interact to speed processing the data significantly. A quantum jump forward of three orders of magnitude is projected, two from the efficiencies and reduced computing load made possible by the MLG algorithm and one from the new parallel strategic computing architectures.

Figure 1 depicts the major elements of the SCAT problem inherent in battlefield management. An array of sensors, indicated as round dots, samples the actual 3D battle volume or "ground truth" in the region of interest at various times "t_{si}" where the index "s" corresponds to a particular sensor and the index i = 1, 2, ..., n, identifies the times at which observations for that particular sensor take place. The observations by sensor

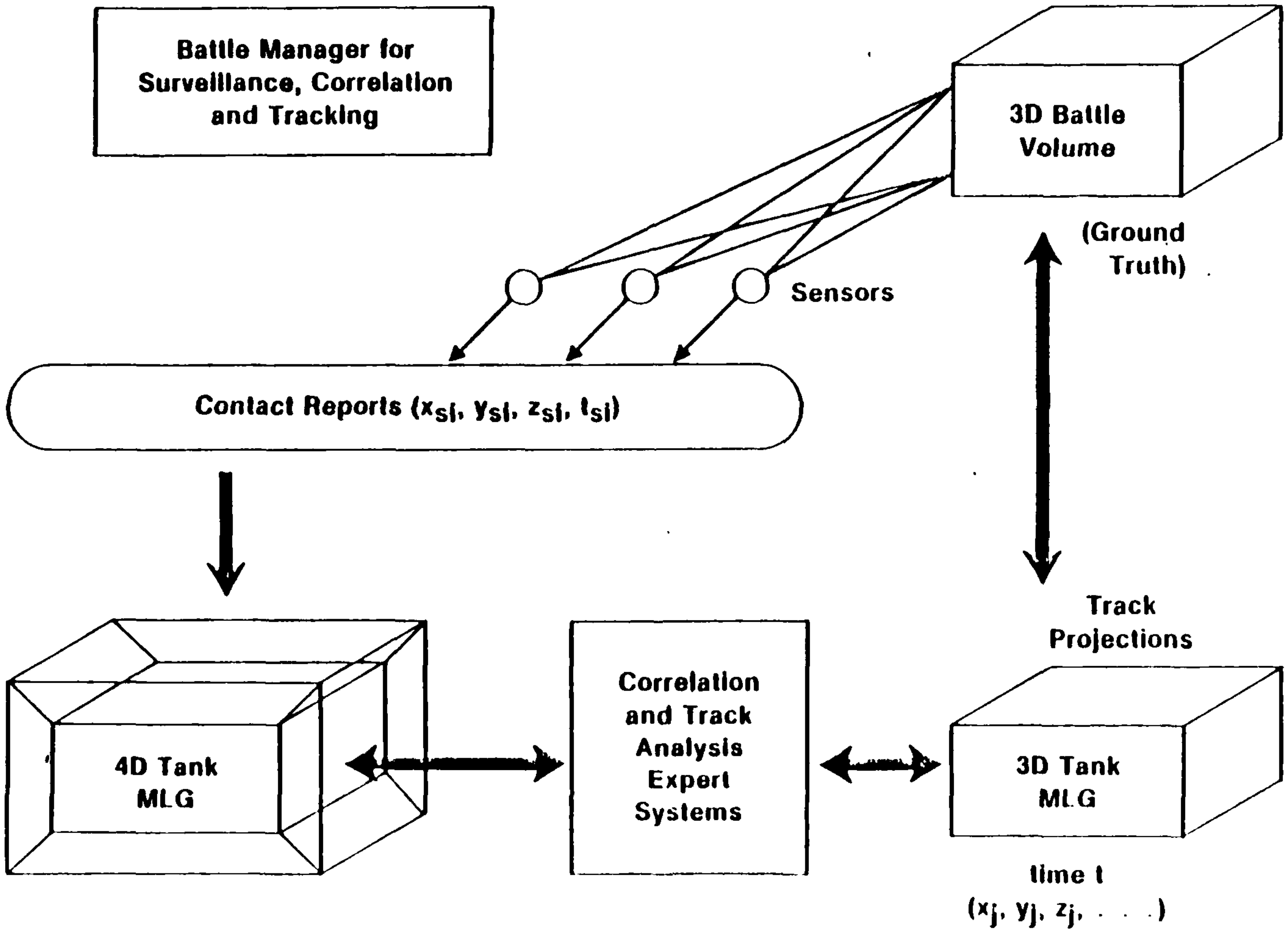

Battle Manager for Surveillance, Correlation and Tracking
3D Battle Volume
(Ground Truth)
Sensors
Contact Reports (x_{si}, y_{si}, z_{si}, t_{si})
Track Projections
4D Tank MLG
Correlation and Track Analysis Expert Systems
3D Tank MLG
time t
(x_j, y_j, z_j, . . .)

Figure 1

s at a time t produce a number of "contact reports," each of which corresponds to specific locations x, y, z, within the battle area. The index "i" labels a particular contact report which has been produced by the sensor "s."

Each contact report possibly corresponds to a prospective target. One can expect no sensor to detect all of the targets which are present in the engagement area, and some number of contact reports will correspond to objects other than targets of interest. In addition, the locations and times specified in each contact report have associated errors or uncertainties which depend on the properties of sensor s, the environmental conditions which the sensor encounters at time t_{si}, and the relative location of the sensor in the battle volume. These errors will be part of the data provided as contact reports and thus will pollute the correlation of specific contact reports with each other as distinct views of the same object provided by the array of sensors being used. The goal is to determine, as accurately as possible, how many things are there and their characteristics, what the detailed spatial configuration of combatants is, and what their velocity or course is at the present time t based on all the observations at previous times (t_{si}). The key performance factors are the frequency of observations by each sensor, the completeness and reliability of the information which each sensor provides on each target, the mathematical, algorithmic, and practical efficiency of the programming, and the time required by computers to process the available information.

As shown in Figure 1, the processing of sensor data occurs in two major computational activities connecting four distinct software and hardware systems. First, the data must be indexed according to times and spatial locations of all observations. That is, the contact report labels x, y, z, t must be translated into indexes i, j, k, ℓ for storage in a four-dimensional "tank." We define a "tank" as an active, random access data structure which is continually restructured to adapt to the changing configuration implied by new contact reports, known combatant motions, or projecting correlated tracks. The 4D tank (an MLG data structure similar to the 3D MLGs used in molecular dynamics by the LCP) evolves primarily in response to new contact reports. The computer system which serves as the hardware portion of the tank will have specific, limited capacity. Thus as new data enter the 4D tank, the oldest data must be dumped into an archival storage system (OATH). The MLG algorithms are ideally suited to this archival requirement because the earliest plane is a 3D MLG which can be moved in bulk.

The second step in processing must involve the cooperation of "expert" systems, pattern recognition, associative memory, and other computational decision aiding or decision making to interpret and evaluate the sorted contact reports in the 4D tank. Again the MLG ensures the instant availability of all the data required for detailed local calculations of a type which will occur with high frequency (e.g., fusing seven radar observations of one target in the presence of additional targets passing at slightly different angles and speeds). The composite analysis of

the geo-temporally possible "interactions" of the contact reports, as efficiently determined by the MLG, determines a projected estimate of the real situation, the "ground truth," at the present time "t." The MLG approach does not say how to do this "thinking" but it does say how to screen out significantly the amount which must be done, it gives an efficient framework to do the screening against, and will provide a networked operating simulation system within which varied methodologies can try their hand at high level screening.

The projected locations of known objects and inferred target tracks, i.e., the "answers," will be entered into another MLG data structure, a 3D surveillance tank, which stores the instantaneous best estimate of what the real configuration is as time passes. "Battle managers" access the reduced data set stored efficiently in the 3D tank to form strategies and to calculate a response to threats. The faster the bulk sorting and preliminary analysis of contact reports takes place, the faster the multisensor fusion will be accomplished and the more recent will be the observations which can contribute to that analysis, and the more accurate will be the estimate of present "ground truth" which resides in the 3D tank.

The analysis of 4D contact report data requires identification of contact reports which are the (geo-temporal) near neighbors of each contact report in the tank. Similarly, battle managers using the 3D tank must have information showing the geographical near neighbors of a given target in order to respond to the threat most effectively. Rapid computing of geo-temporal or

geographical proximity is hampered by the poor N^2 scaling inherent in the need to determine $N(N - 1)/2$ distances between a total of N contact reports (4D tank) or N projected target locations (3D tank). For large numbers of targets this represents a combinatorial explosion which must be contained from the very onset of the battle management activity. The MLG algorithm accomplishes the containment in three ways:

1) The MLG automatically orders the 4D and 3D data structures so that adjacent indices in the data structure must correspond to contact reports and target projections, respectively, which are near neighbors. Data access is then the most rapid possible on a given computer system.

2) The speed at which the MLG data structures are determined scales as N log N, so that the calculation of the near neighbors of any contact report (4D tank) or target projection (3D tank) is N/(log N) times as fast as by other methods. This amounts to an improvement of more than two orders of magnitude in processing speed for the number of targets being considered by SDI.

3) The MLG algorithm seems ideally suited to future computer systems which rely on parallel processing using both distributed control with asynchronous processing and localized control with synchronized processing.

Figure 2 describes more specifically the MLG simulator toward which this proposal is directed. The simulator makes substantial

/2C

Battle Engagement Area Surveillance Tank

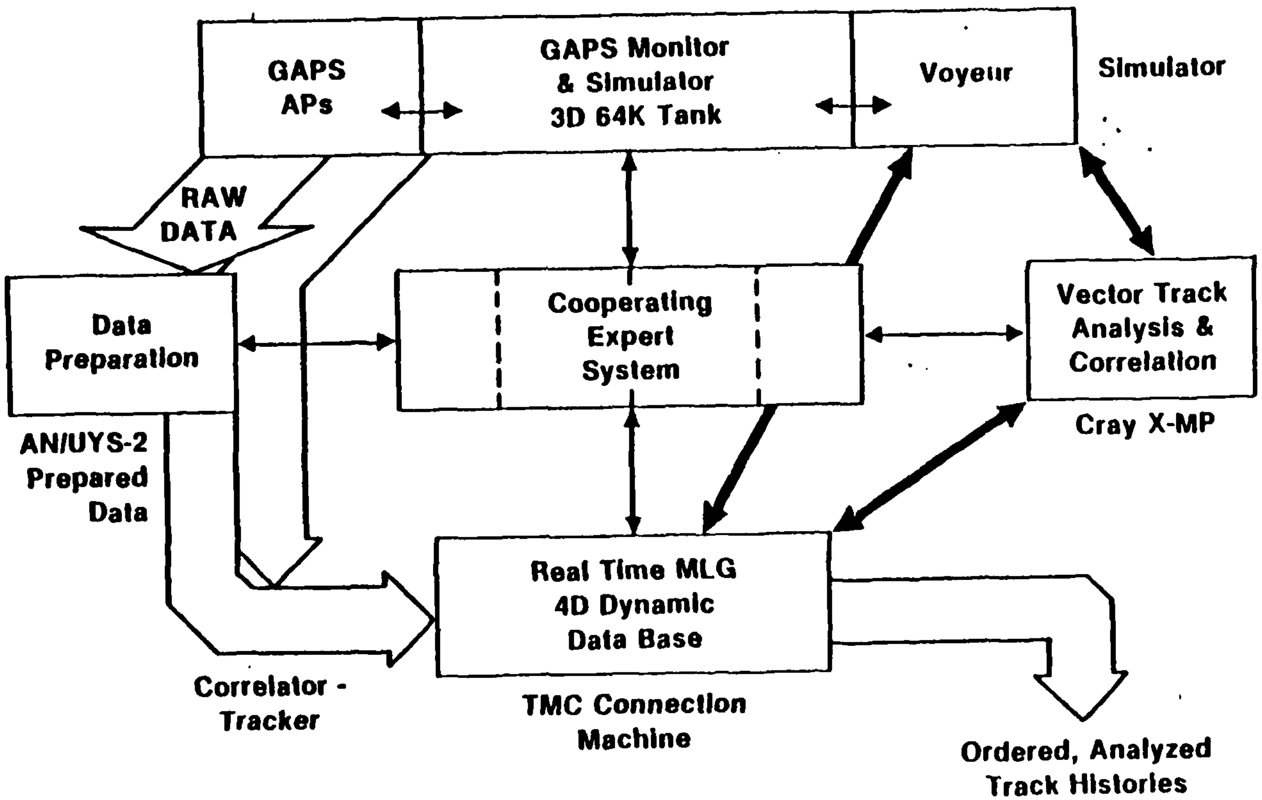


Figure 2

use of equipment which has already been acquired by the Laboratory for Computational Physics (LCP) or the Naval Research Laboratory (NRL) and equipment which is deliverable to NRL in accordance with other funding. The elements of the system are as follows:

1) The LCP GAPS provides a 3D tank for the real-time representation and modeling of ground truth and would also provide the simulated raw sensor data for incorporation into the 4D tank described above. This capability thus serves to define the engagement area and would be a central software component in all engagement area simulations and real-time analyzers.

2) The Battle Engagement Area Surveillance Tank (BEAST) project would develop a 4D MLG for sorting and organizing incoming contact reports in real time. The structure of the MLG algorithm ensures the user that each contact report is stored logically adjacent to the other reports which are nearest it in space and time in the battle area volume. This allows the most rapid analysis and correlation of data by the expert systems. We anticipate that the DARPA Thinking Machines Corporation (TMC) Connection Machine (CM), which will arrive by early 1986, will ultimately serve as the 4D tank.

3) The NRL Cray X-MP will perform the track analysis and correlation functions along with other networked systems. The interaction between the Cray and other (expert) systems co-resident in the NRL network is

depicted in Figure 2. The results would be stored and displayed through a 3D MLG surveillance tank, also in the Cray. Specifically, the 3D tank would contain the projected battlefield situation for the present time "t." The "Battle Manager" would access this information in several ways. We currently envision using graphical systems based on the asynchronous, interactive software package VOYEUR. Such a display system is currently operational on the LCP GAPS, using multiple array processors executing concurrently on a complex multidimensional fluid simulation.

More detailed descriptions of the major elements of the MLG approach and restructuring algorithms appear in the following references.

"A Vectorized 'Neighbors' Algorithm of Order N Using a Monotonic Logical Grid," by J. P. Boris, presented at the 26th APS PPD, October 29-November 2, 1984, Boston. Bull. Am. Phys. Soc. 29 (5), Paper 4F9, 1262 (October 1984).

"A Vectorized 'Nearest-Neighbors' Algorithm of Order N Using a Monotonic Logical Grid," by J. P. Boris, NRL Memorandum Report 5570 (May 1985).

"A Vectorized 'Nearest-Neighbors' Algorithm of Order N Using a Monotonic Logical Grid," by J. P. Boris (to be published in J. Comp. Phys.).

"A Vectorized Near Neighbors Algorithm for Organizing Free Lagrangian Dynamics Models," by Jay P. Boris and S. G. Lambrakos, Proceedings of The Free Lagrangian Dynamics Conference, 4-6 March 1985, Hilton Head Island, South Carolina (to be published by Springer-Verlag).

"Dynamical Organization of Evolving Lagrangian Data Using a Monotonic Logical Grid," by Jay P. Boris and S. G. Lambrakos, Proc. of 1985 Summer Computer Simulation Conference, 22-26 July 1985, Chicago.

"A Vector Near Neighbors Algorithm for Organizaing Free Lagrangian Fluid Dynamics Models Using a Monotonic Logical Grid," by Jay P. Boris, Proc. of the International Symposium on Computational Fluid Dynamics, 9-12 September 1985, Kenchiku Kaikan, Tokyo.

"Geometric and Statistical Properties of the Monotonic Logical Grid Algorithm for Near Neighbor Calculations," by S. G. Lambrakos and J. P. Boris (submitted to J. Comp. Phys.).

"A Vector Algorithm of Order N Solving the Near Neighbors Problem in Manybody Particle Dynamics," by S. G. Lambrakos and J. P. Boris, Eleventh International Conference on Numerical Simulation in Plasmas, Montreal, Canada, June 24-27, 1985.

"A Highly Parallel Near Neighbors Algorithm of Order N Based on a Monotonic Logical Grid," by S. G. Lambrakos and J. P. Boris, Second SIAM Conference on Parallel Processing for Scientific Computing, November 18-21, 1985, Montreal, Canada (to be published in the Pmoceedings).

"A Vectorized Near Neighbors Algorithm of Order N for Molecular Dynamics Simulations," by S. G. Lambrakos, J. P. Boris, I. Chandrasekhar, and B. Gabor, International Symposium on Computing in Chemistry, New York Academy of Sciences, October 2-4, 1985.

APPENDIX B

A Vectorized "Nearest-Neighbors" Algorithm of Order N Using A Monotonic Logical Grid

by

Jay Boris

Laboratory for Computational Physics

U.S. Naval Research Laboratory, Code 4040

Washington, DC 20375

Accepted for publication in:

The Journal of Computational Physics

TABLE OF CONTENTS

126

ABSTRACT

When a large number of separate objects interact, N*(N-1)/2 interactions can occur. At any instant a given object may interact strongly with only a few of the N - 1 others. Unfortunately, keeping lists of the other objects with which it interacts or recomputing these nearest neighbors each timestep is computationally expensive. This "nearest-neighbors" problem has persisted in computational physics and computational geometry for several decades. We need efficient algorithms which select important nearest-neighbor interactions without having to check and analyze N^2 interactions. To date the best algorithms which scale as N, rather than N^2, are scalar algorithms which address memory randomly.

This report introduces an efficient 3D nearest-neighbors algorithm whose cost scales as N and which vectorizes easily using data from contiguous memory locations. A Monotonic Logical Grid (MLG) for storing the object data is defined dynamically so that objects which are adjacent in real space automatically have close address indices in the compact MLG data arrays. The data values for each object are stored at a location (i,j,k) in the MLG such that the X positions of all the objects increase monotonically with index i, the Y positions increase monotonically with index j, and the Z positions increase monotonically with index k. Such a well-structured mapping from the real positions to regular, compact data arrays can always be found. Further, when object motions result in a local violation of spatial monotonicity, another MLG always can be found nearby. This means that local changes in the object positions and hence spatial ordering do not trigger global changes in where these object data are stored in the MLG.

The data relocations required to maintain the MLG as objects pass each other in space can also be vectorized efficiently. The MLG algorithms will execute effectively in small array processors and partition to take advantage of asynchronous parallel architectures in VLSI/VHSIC-based supercomputer systems of the future. Using a commercially available distributed processing system, 5000 interacting objects could be monitored and the MLG data base updated and restructured thousands of times in about 15 minutes, fast enough to be useful for real time applications as well as physics simulations.

SECTION I. INTRODUCTION AND BACKGROUND

When N independent objects interact in space, $N(N-1)/2$ interactions might be important in determining how a given object reacts to the others at any instant. Usually exact positions and velocities of the neighboring objects must be known. Knowing statistical averages and the general properties of the distribution of objects nearby does not provide enough data to compute local interactions accurately. At any instant a given object may interact strongly with only a few of the others. Unfortunately, keeping track of or repeatedly recomputing which ones are near neighbors is computationally expensive. The goal is efficient, simple algorithms which select the nearest-neighbors without a computational premium scaling as N^2. Effort on the nearest-neighbors problem has persisted in computational physics and computational geometry for several decades. This report introduces an efficient 3D nearest-neighbors algorithm whose cost scales as N and which vectorizes easily using contiguous memory locations.

An efficient vector solution of the nearest-neighbors problem would advance many important applications. For an important class of molecular dynamics problems involving interactions among many atoms and molecules, the nearest neighbors exert the strongest forces and are the most likely candidates to enter into chemical reactions. Many important physics problems in gases, liquids, solids, and transitions among these phases require detailed manybody calculations where the close encounters are most important.

For graphics based on vertex-edge representations of complex 3D shapes, local relationships and orientations of nearby vertices determine which surfaces are visible. It is clearly advantageous to be able to construct a 2D image of a complex 3D scene, for example, using the parallelism made possible by Very Large-Scale Integration (VLSI). Terrain management simulation models and multi-dimensional radiation transport models are currently limited in their ability to compute geometric obscuration. For controlling airline traffic over crowded airports, collisions with nearby planes are the most immediate danger - and demand shorter timescales for detection and corrective response. These applications all require rapidly updating many distinct local configurations as the objects move.

For complex manybody problems with N = 5000 independent objects, more than thirty thousand of degrees of freedom are required, and 12.5 million interactions exist which ideally ought to be considered. Current supercomputers deliver ~50 Megaflops (million floating point operations per second) on optimized but realistic problems. The straightforward recalculation of all interactions requires about 60 vectorizeable operations per interaction, or 10 - 15 seconds of dedicated supercomputer time. This is not fast enough for real time applications where the data base should be updated and the neighbors recalculated every second or two.

Note added in proof: A second report by Lambrakos and Boris has been prepared on the geometrical properties of the MLG method in which statistics and performance are discussed. This report has been submitted to the Journal of Computational Physics and will appear as an NRL memorandum report. The technique is being used in molecular dynamics calculations in three dimensions.

This report introduces a simple three-dimensional nearest-neighbors algorithm whose cost scales as N, not as the square of N, and which vectorizes easily using data from contiguous memory locations. A compact data structure to store the object data, called a Monotonic Logical Grid (MLG), is defined dynamically so that objects which are adjacent in real space automatically have close address indices in the MLG data arrays as well. As two objects move past each other in space, their data are exchanged or "swapped" in the MLG data arrays to keep a strictly monotone mapping between the geometric locations and the corresponding storage location indices.

To construct an MLG the data values for each object are stored at location (i,j,k) in the MLG such that the X positions of all the objects increase monotonically with index i, the Y positions increase monotonically with index j, and the Z positions increase monotonically with index k. Section III describes the algorithm in some detail.

It is not obvious but it is true that such an organized logical ordering of even random locations can always be found. In Section III an order N log N constructive algorithm for one such MLG is provided proving existence. Generally more than one MLG meeting all the monotonicity conditions seems to be possible so the technical problem of selecting the optimum MLG for a particular application has to be addressed. In one case, minimizing average distances to neighbors in the MLG may give the best grid. In other problems it may be best to maximize the shortest distance to any point which is not a near neighbor in the logical grid.

Further, when object motions result in a local violation of the monotonicity conditions on which the original MLG was based, another MLG can be found nearby. This means that local changes in the object positions and hence spatial ordering do not trigger global changes in where these object data have to be stored in the MLG. The data relocations to maintain the MLG as objects pass each other in space can be vectorized without inefficient gather/scatter operations or variable-length (scalar) linked lists. The MLG data structure and algorithms allow contiguous-data vector operations which are long enough to be efficient for physical force sums, for $F = Ma$ orbit integrations, and for the object data "swapping" used to restructure the MLG whenever the monotonicity conditions are violated.

The cost to execute a simple test version of the model is one hour on a DEC VAX 11/780 for one thousand particles for 1000 timesteps. We used a power series force law for the 124 nearest neighbors, assuming that the average particle separation distance is smaller than the cutoff radius R_c of the force law. A commercial distributed processing system, a minicomputer host with modest array processors, would be fast enough using an MLG to integrate 5000 interacting objects and restructure the data base thousands of times in about 15 minutes, useful for realtime applications where current supercomputers using other algorithms will be inadequate. The MLG also permits partitioning to take advantage of asynchronous multiprocessor parallelism in VLSI/VHSIC-based distributed processing systems (e.g. Fox and Otto, 1984).

The computational costs of several algorithms and combinations of algorithms which have been proposed to reduce the cost of finding nearest neighbors are discussed in Section II. Section III contains a description

of the MLG itself, an O(N log N) sort algorithm to find a starting MLG from arbitrary initial data, and simple algorithms which restructure the grid dynamically as the objects move. Section III also presents a few simple tests of the method. Section IV considers several extensions. Section V contains a summary and conclusions.

SECTION II. THE COMPUTATIONAL COST OF FINDING NEAREST NEIGHBORS

As described in Section I, brute force recalculation of all the interactions can be vectorized but is of order N^2 and therefore costly. The best nearest-neighbors algorithms published, Hockney and Eastwood (1981), are of order N with minimal operation counts. However, these O(N) algorithms are intrinsically scalar and execute relatively poorly in parallel or pipeline-architectured supercomputers. Further, memory is addressed essentially at random so data buffering from disk or virtual memory for a large problem is time consuming.

In the next few paragraphs the computational costs of simple strategies to reduce the nearest-neighbors problem are compared. The assumed operation counts are only representative values since optimal implementations are always machine and problem dependent (Gunsteren, et al., 1984). There are too many possible algorithms and variations to compare all of their operation counts. It is even harder to compare scalar and vector algorithms, something we would like to do in theory here but which can really only be done in practice. As a rule of thumb, efficient use of the vector hardware in supercomputers or array processors generally produces over an order of magnitude speed up over reasonably well compiled scalar code. In some cases the vector speed-up factor will be greater and in others, less.

For problems where the number of near neighbors is large so the computational cost is dominated by the physical interaction calculations rather than computational book-keeping to find the near neighbors, the algorithm introduced here calculates two to three times as many interactions as minimally necessary. This is accepted as the price for simple logic

and vectorized computation in contiguous memory. This means that a computer whose vector speed is only a few times the scalar speed may see no improvement over the Hockney-Eastwood PPPM techniques. In computers where the vector-scalar ratio is large, an order of magnitude improvement with an MLG is at least conceivable. More substantial gains are possible in highly parallel multi-processor systems because the MLG algorithms partition naturally.

Let $N_{ot} = {}^{\sim}60$ be the total number of floating point operations (flops) used to evaluate each interaction between two of the N = 5000 objects. The main component of the cost for a timestep will be

$$\begin{aligned} \#\ \text{Flops to compute all interactions} &= \text{Fcai} \\ &= N \times (N \times N_{ot})/2 \qquad (2.1) \\ &= 7.5 \times 10^8 \text{ flops} \rightarrow 15 \text{ seconds/step at 50 Megaflops.} \end{aligned}$$

Manybody calculations which compute all interactions have become prohibitively expensive with even a few hundred objects because thousands of timesteps are required for complex problems. The operation count per timestep goes up quadratically with the number of particles N but the effective resolution only increases as the cube root of the number of particles. This scaling of cost with at least the sixth power of resolution is prohibitive. If the number of timesteps also has to be increased when more objects are simulated, the scaling can be even worse. This brute force $O(N^2)$ algorithm is of interest because it vectorizes and partitions easily and is exceedingly simple.

Reduction of this computational expense is obtained by computing the details of the interactions only for pairs of objects closer than a cutoff distance R_c. This basic nearest-neighbors concept takes its most

sophisticated form in the "Particle-Particle-Particle-Mesh (PPPM) algorithms of Hockney and Eastwood (1981). Faster algorithms and data structures for implementing this nearest-neighbor approximation have been the subject of much computational research in the last few decades.

Checking two locations to see if they are within a distance R_c of each other requires about $N_{oc} = \sim 10$ floating point operations. Nine or ten flops are required simply to calculate the square of the distance between the two objects and then compare it with the square of R_c.

To calculate the physical forces and sum them for each interaction pair separated by less than R_c requires

$$N_{oi} = N_{ot} - N_{oc} = \sim 50 \text{ floating point operations} \qquad (2.2)$$

per interaction. This would be enough operations to calculate, for example, a simple force law with transcendental functions or to estimate a collision cross-section. If $N_{nn} = \sim 60$ neighbors must be considered for each object, our formula for the number of flops to compute near interactions is

$$\begin{aligned} \#\ \text{Flops to compute near interations} &= \text{Fcni} \\ &= N \times (N/2 \times N_{oc} + N_{nn} \times N_{oi}) \qquad (2.3) \\ &= 1.4 \times 10^8 \text{ flops} \rightarrow 2.8 \text{ seconds/step at 50 Megaflops.} \end{aligned}$$

Using the concept of a cutoff radius R_c has reduced the operation count by over a factor of five. Just checking all the interaction distances requires appreciable time when $N = 5000$ and N_{oc} is a modest fraction of

N_{oi}. The computational cost still scales as N^2 and the factor of five speedup shown above is largely illusory as it ignores the cost of organizing the neighboring object data into contiguous vectors for efficient computation of the physical interactions. If this were not done, the interaction calculations would have to be performed as scalar operations. It would be worth organizing these vectors if enough objects interact with each other on average but this further optimi-zation would be wasted if the average particle separation distance were bigger than R_c.

In practice the N_{nn} near neighbor variables have to be gathered into vectors, costing about $N_{gv} = \tilde{}\ 50$ flops per neighbor per object. The vector lengths are also shorter, engendering some additional overhead. Thus a more accurate estimate of F_{cni} is

$$\begin{aligned}\text{\# Flops to compute near interactions} &= F_{cni} \\ &= N \times (N/2 \times N_{oc} + N_{nn} \times (N_{gv} + N_{oi})) \\ &= 1.75 \times 10^8 \text{ flops} \rightarrow 3.5 \text{ seconds/step at 50 megaflops.}\end{aligned} \tag{2.4}$$

Reducing the number of nearest neighbors used does not help much as long as cutoff radius checking is being done for all possible inter-actions. In fact, as long as this is done, it hardly hurts to double the number of neighbors kept for calculating the interactions. Clearly, too much time is spent checking interaction distances which are too large to be significant.

It is natural to consider updating the lists of near neighbors less often. If the $O(N^2)$ selection of near neighbors is only done every $N_{sc} = \tilde{}2.5$ cycles, the cost for this portion of the algorithm is reduced. However, there are additional costs and N_{sc} cannot be very

large. The number of near neighbors has to be augmented to allow for motion of particles near R_c into and out of the cutoff region. This increase is roughly the same as the relative volume change due to particles moving into and out of the sphere of radius R_c during the N_{sc} steps between recomputation of the near neighbors lists. Let an average particle travel a small fraction f of R_c per timestep. The sphere of near neighbors will increase in volume by a factor of $\sim(1 + f \times N_{sc})^3$ in the N_{sc} steps until the neighbor lists are updated. When the sphere has roughly doubled in volume, a radial change of 25%, the lists must be updated since most of the neighbors will have changed. Thus, with $f = 0.1$

$$N_{sc} = .25/f = \sim 2.5 \tag{2.5}$$

and N_{nn} must be multiplied by $(1 + f \times N_{sc})^3 = \sim 2$ in the above formula to give

$$\begin{aligned} \text{\# Flops to intermittently recalculate neighbors} &= F_{irn} \\ &= N \times (N/2 \times N_{oc}/N_{sc} + N_{nn} \times 2 \times (N_{gv} + N_{oi})) \qquad (2.6) \\ &= 1.1 \times 10^8 \text{ flops} \rightarrow 2.2 \text{ seconds/step at 50 megaflops.} \end{aligned}$$

Although the N^2 search is a factor of 2.5 less important, counteracting effects eat up some of this gain. If f were reduced to 0.01, probably reasonable for molecular dynamics simulations, some additional gains could be realized. However, these would be better by less than a factor of two

because the cost is now in computing the near interactions as it should be. In other applications, however, there would not necessarily be a need to use the short timesteps required by physical simulations such as molecular dynamics.

The only way to avoid the N^2 premium is to update the nearest-neighbors list of each object using objects in a volume larger than would be required for an interaction cutoff of R_c but much smaller than the entire system. Hockney and Eastwood define a PPPM "chaining mesh" where $dX = dY = dZ = R_c$ and check distances to objects known to be in only the nearest 13 $= (3^3 - 1)/2$ cells. Only these particles might be within R_c of a particle in the chaining cell under consideration. On average only about 40% of the particles in these 13 cells are actually within R_c. Taking L as a typical system dimension, there are $N_{cc} = N\ (R_c/L)^3$ particles in each of the PPPM cells. The number of cutoff distance checks performed in a timestep is then

$$\#\ \text{PPPM checks} = N \times 13 \times N_{cc}. \qquad (2.7)$$

In the PPPM formulation, when R_c is twice the average spacing, a typical particle has its distance to 104 other nearby particles checked (13 cells × 8 particles per cell). The corresponding number in the MLG would be 62 if all interactions within two grid displacements in any direction are kept. This nominal factor of two gain in the MLG approach is lost again because all the interactions would be calculated to maintain vectorization rather than only 40% as possible with the scalar PPPM algorithm. The real gain is the ability to use efficient, contiguous memory, vector operations

throughout the MLG algorithms and to cleanly partition the problem into computational subtasks.

The operation count for the overall MLG algorithm developed next in Section III is also problem dependent. Using typical simulation variables summarized in Table 2.1 below, the cost of the MLG in vector floating-point operations to execute a timestep, exclusive of the relatively inexpensive orbit calculations, is

$$\begin{aligned}
&\text{\# Flops for the Monotonic Logical Grid algorithm} = F_{mlg} \\
&\quad = N \times (N_{nn} \times N_{ot} \quad \text{for neighboring object interactions} \qquad (2.8) \\
&\quad\quad + 3 \times N_{si} \times N_{os}) \quad \text{for swapping iterations in X, Y, Z} \\
&\quad = 2.25 \times 10^7 \text{ flops} \rightarrow \sim 0.5 \text{ seconds/step at 50 megaflops.}
\end{aligned}$$

Here $N_{si} = \sim 4$ is the number of iterations of vector swapping performed over the entire grid to restructure the MLG after the object positions change each timestep. $N_{os} = \sim 60$ is the number of floating point operations to execute a single swap of two objects in the MLG. In the next section the concepts and details of the Monotonic Logical Grid Algorithm are presented.

/40

Table 2.1. Typical Values of Manybody Simulation Variables

N	=	5000	=	# of objects interacting in space
N_{os}	=	~60	=	# operations per vector swap in MLG algorithms
N_{si}	=	<4	=	average # of vectorized swapping iterations to relocate object data in the MLG
N_{nn}	=	~60	=	# of near neighbors usually included in the interaction calculations
N_{oc}	=	~10	=	# of flops to check distance between two objects
N_{oi}	=	~50	=	# of flops to interact two objects
N_{ot}	=	~60	=	# of flops total to compute an interaction
N_{gv}	=	~50	=	equivalent # of arithmetic flops to gather neighboring object data from random locations in memory
N_{sc}	=	~2.5	=	# of steps between recomputation of the nearest-neighbors lists in scalar algorithms
N_{cc}	=			# of objects in the average cell of PPPM chaining mesh.

III. THE MONOTONIC LOGICAL GRID ALGORITHM

A Monotonic Logical Grid (MLG) is a simple, compact way of indexing and storing the data describing a number of objects moving in space. For N particles in three dimensions, the arrays of object locations, X(i,j,k), Y(i,j,k), and Z(i,j,k), constitute an MLG if and only if

$$\begin{aligned} X(i,j,k) &\leq X(i+1,j,k) \quad \text{for} \quad 1 \leq i \leq NX-1, \\ Y(i,j,k) &\leq Y(i,j+1,k) \quad \text{for} \quad 1 \leq j \leq NY-1, \text{ and} \\ Z(i,j,k) &\leq Z(i,j,k+1) \quad \text{for} \quad 1 \leq k \leq NZ-1. \end{aligned} \tag{3.1}$$

Given N = NX*NY*NZ random locations, the spatial lattice defined by an MLG is irregular. However, the cells defined by logically neighboring locations are distorted cubes and thus form a useful consistent partitioning of the spatial volume. When the N object locations satisfy Eqs. (3.1) and any additional constraints or relations specifying other than infinite-space boundary conditions, they are in "MLG order". This ordering is useful because the direction for going from one object to another in space and in the MLG are the same. Further, other objects which are between two objects in space will also be between them in the MLG. Thus neighbors in real space have neighboring address indices in the MLG as well.

Figure 3.1 shows three different spatial configurations of 16 objects in the lower three panels. The objects are ordered into four rows and four columns in each of these configurations corresponding to regular storage of the object data in the two-dimensional MLG shown above. The cells of the MLG move with the objects and thus always have exactly one object in them. When all the objects move to the upper left of the region, as in the lower right panel, the MLG is just as regular as when the objects

are uniformly spaced. This mapping of irregular locations onto a very regular data structure is what permits optimal use of vector and multiprocessor hardware.

Figure 3.2 illustrates several different MLG mappings of the same 16 object locations. The upper left panel shows the 16 locations in a regular spatial lattice. The obvious numbering of the locations into four rows of four objects each is an MLG because the X and Y components of all the object locations (dots) increase monotonically with the X and Y indices, i and j. In the remaining three lower panels the locations have been displaced from the regular spacing. Each of these panels contains the same data, but the MLGs for storing these data, as indicated by the logical mesh lines, all differ. The lower left panel is a recognizeable distortion of the regular grid above. The indexing of the objects would be identical in both cases although the actual data stored would differ somewhat since the objects have moved away from their regular locations. The lower center and lower right panels show different logical indexing for the same physical data giving two other Monotonic Logical Grids (MLGs). In the center the connections to points in the second row from above and from below have all been displaced to the left. On the right, the connections to the second column from objects located logically in columns one and three have been displaced downward.

These figures show that there can be a number of MLGs with the same Lagrangian object data, all satisfying the required monotonicity conditions from Eqs. (3.1). These spatial monotonicity conditions constitute 3N - NXxNY - NYxNZ - NZxNX numerical comparisons which can be performed to determine if a particular organization of the object locations is in MLG

order. For each dimension of the desired data structure such a monotonicity condition can be defined.

In space the coordinates can be rotated or redefined and this corresponds to a different family of MLGs. The monotonicity conditions may not change in the new coordinate system but the object location coordinates will. Even if the coordinate system is held fixed, the object motions will quickly invalidate at least some of the relations (3.1) requiring a reorganization of the object data in the arrays to store a completely monotone mapping. Using the monotonicity conditions, a given data structure can be checked efficiently to see if the locations are in MLG order. However, additional algorithms are needed when MLG order is violated.

If the objects are not in MLG order, the following algorithm using a vector sort routine O(N log N) can be used to rearrange them. First sort all N locations into the order of increasing Z. The first NX x NY of them, should be indexed k = 1, and sorted into the order of increasing Y. The first NX of these, should be indexed j = 1 and then sorted into the order of increasing X. These objects are indexed from i = 1 to i = NX. The next NX locations, indexed j = 2 but still k = 1, are again ordered and indexed from i = 1 to i = NX. This procedure is continued until the first NX x NY plane of locations has been arranged. Since the locations were initially ordered in Z, the subsequent reorderings within the k = 1 plane cannot disturb the monotonicity conditions relating the first plane to any subsequent reordering of the second and subsequent planes. Similarly, all the locations will satisfy the monotonicity conditions in Y and X as well.

Once the first plane is ordered, the next NX x NY locations are indexed k = 2, and the MLG ordering within this plane is constructed just as for the first plane. All NZ planes are organized this way. The process requires of order

$$\begin{aligned}
&NZ \times NY \times NX \times (\log NZ + \log NY + \log NX) \\
&\quad + NZ \times (NY \times NX \times (\log NY + \log NX) + NY \times NX \times \log NX) \qquad (3.2) \\
&\quad = NZ \times NY \times NX \times (\log NZ + 2 \log NY + 3 \log NX)
\end{aligned}$$

operations to construct the MLG. This sort algorithm could be repeated every timestep as necessary to restructure the MLG when object motions in one of the three coordinate directions cause some of the conditions (3.1) to be violated.

The existence of this constructive algorithm proves that at least one MLG for even random locations always exists and that it is not hard to find. As a consequence, data manipulation and summation algorithms in the MLG can always assume the rigorous spatial monotonicity of the MLG. When several object locations are identical, any ordering the sort procedure comes up with is correct as the conditions (3.1) are satisfied. Locally degenerate grids are possible when several locations overlap.

Although this algorithm is fast, it has two limitations: it is of order N log N, not N, and it may move data a long distance in index space to correct even small changes in position. To counter these objections, an order N algorithm is described which executes local but vectorizeable exchange or "swapping" operations on the MLG data to restore monotonicity everywhere. The extra factor log N is removed because small monotonicity upsets from the previous set of locations generally do not require information from the other side of the grid for their correction.

If two objects move less than a typical separation distance per timestep, a condition generally required for accurate integration of the equations of motion, a few iterations are usually enough to restore MLG order. A "swap" is executed by testing the conditions in Eqs. (3.1), and

then, when the corresponding monotonicity condition is violated, exchanging the locations in the logical grid of all data pertaining to the two objects involved. Each direction is checked separately. A red-black algorithm (Adams and Jordan, 1984) would allow at least half the tests in a given direction to be performed simultaneously and thus vectorized while converging as fast as a scalar iteration.

No more than five arithmetic operations are required to test for monotonicity and to prepare to swap any amount of data. A floating point 0.5 is logically "or"ed with the sign bit of the coordinate difference to obtain a number S which is 0.5 if the two coordinates are in MLG order and -0.5 if they are not. This takes three operations. Two more operations give the weights w and (1 - w) where

$$w = S + 0.5, \qquad (1 - w) = S - 0.5. \tag{3.3}$$

The weight w is unity when the coordinates are in order and zero when they are not. If the test $X(i,j,k) \leq X(i+1,j,k)$ is being performed, the X components of the object locations can be exchanged using the following formulae (six operations):

$$\begin{aligned} T(i,j,k) &= w \times X(i,j,k), \\ U(i,j,k) &= (1 - w) \times X(i,j,k), \\ X(i,j,k) &= T(i,j,k) + (1 - w) \times X(i+1,j,k), \text{ and} \\ X(i+1,j,k) &= w \times X(i+1,j,k) + U(i,j,k). \end{aligned} \tag{3.4}$$

When the locations are in MLG order, the swapping formulae change nothing. When two locations are out of order, these formulae interchange the object data on the MLG so they will be in order for the next iteration. All object data at every grid point can be treated identically with this procedure.

The algorithm vectorizes easily even though some of the pairs of objects have to trade places in the MLG and others do not.

These six operations must be repeated to swap each data variable stored in the MLG. As a minimum these include the three components of the object locations and an identification number, ID#(i,j,k), to mark which of the N particles currently is at i,j,k in the MLG. To vectorize the complete algorithm, the velocity components VX(i,j,k), VY(i,j,k), VZ(i,j,k), the mass M(i,j,k), and another force law constant FC(i,j,k) must also be moved about dynamically. These nine variables require 54 operations to be moved between adjacent cells for each swapping iteration. Thus Nos = ~60 operations are required for each iteration in each direction for each object. This is about as much work as calculating three components of the force acting between two objects which are near neighbors in the MLG. With Nsi = ~4 swapping iterations being performed in each direction, the total cost of restoring the MLG every timestep is about the same as calculating forces from 12 neighbors. When timesteps are short, this can be reduced even further.

When the MLG algorithm is used, the cost in vector floating-point operations to execute the geometric and force summing in a timestep is given by Eq. (2.8). The speedup expected using this algorithm is large, a factor greater than thirty for 5000 objects. Not only is the N^2 dependence removed but the actual nearest-neighbor interactions can be computed with very high efficiency, comparable to the best order-N scalar algorithms. Only about a fifth of the computation is expended on maintaining the MLG data structure. The rest is used in computing pairs of interactions at full vector efficiency.

The random motion of point particles in a cubical domain is taken as a test problem to illustrate the concepts. A topologically regular $8 \times 8 \times 8$ 3D grid is defined for storing the position and velocity components of 512 randomly located objects. The domain is doubly periodic in X and Y and is bounded in Z by two reflecting end walls at $Z = 0$ cm and $Z = Z_{max} = 8$ dZ. A number of short calculations have been performed using this system to test and develop various aspects of the model. Figure 3.3 shows the first of eight planes of this 3D MLG, plotting the X and Y locations of the 64 objects currently on that plane. The initial conditions for the calculation are shown in the upper right, regularly-spaced locations with random velocities uniformly distributed in each coordinate from -10^7 cm/sec to $+10^7$ cm/sec. The three remaining panels show plots of the 64 locations in the same MLG data plane at three times. As the objects move in the plane and between planes, a complicated but clearly structured MLG is always maintained.

Under a number of different physical circumstances and numerous different initial conditions the model has been able to find an MLG after only a few swapping iterations. The average near-neighbor separations increase somewhat at first over their almost minimal initial values. Rather quickly, however, random swapping halts the increase of this average distance to the near neighbors. Figure 3.4 displays the frequency distribution for the number of swapping iterations required to restore the MLG after relative motion of the objects has disrupted it. Three cases were run from the same physical initial conditions and zero-sized non-colliding particles, with timesteps $\delta t = 2.5 \times 10^{-16}$ sec, 1.0×10^{-15} sec, and

4.0×10^{-15} sec. The lightly shaded bars in Figure 3.4 correspond to the intermediate case with 10^{-15} sec as the timestep. For this case $dl_{max} = .1\Delta$ meaning that the fastest particle traverse 1/10 of the regular initial spacing of $\Delta = 10^{-7}$ cm per timestep. The data with unshaded bars, $dl_{max} = .025\ \Delta$, shows the results when δt is smaller by a factor of four and the data depicted with dark bars shows results when δt is a factor of four larger, i.e. $dl_{max} = .4\Delta$.

To interpret the figure consider $dl_{max} = 0.1\Delta$. About 40% of the timesteps (frequency 0.38) required 4 iterations of swapping to restore the MLG. Less than 10% of the timesteps required 6 or more iterations. The average number of iterations required is 4.0 for $dl_{max} = 0.1\Delta$. When $dl_{max} = 0.025\Delta$, the average number of swapping iterations is 2.85, about $2\sqrt{2}$. When $dl_{max} = 0.4\ \Delta$, the average is 5.0 swapping iterations per timestep. Thus the actual computational work decreases per unit integration time with longer timesteps because the number of swapping iterations increases much more slowly than the timestep increases.

A great deal of swapping goes on in the first few iterations out to the average number for the particular timestep chosen. For timesteps with relatively large numbers of iterations I, the likelihood of this extra work being required decreases by a factor of two or three for each extra iteration. These timesteps requiring a relatively large amount of work contribute very little to the average computation load needed to restore the MLG because they occur so infrequently.

In test calculations, with nonzero particle size, forces were calculated between a given object and the 5 × 5 × 5 cubical nearest neighbors interaction template of 125 neighboring objects centered on it in the MLG. Since the interaction has to be computed only once for a pair of objects and can be ignored for self interactions, the tests had the following number of near neighbors

$$N_{nn} = (5 \times 5 \times 5 - 1)/2 = 62 \approx 60. \quad (3.6)$$

When many objects are within the cutoff distance R_c, the interaction template should be extended, perhaps to 7 × 7 × 7. An appreciable fraction of the forces calculated will be beyond the cutoff distance but this extra work is compensated by the fact that all the work can now be performed by vector operations working from contiguous locations in the computer storage. This gain is typically an order of magnitude or more in speed and is still worthwhile even if a factor of two or three is wasted calculating unnecessary interactions.

When objects are far apart compared to the cutoff radius R_c, only the 13 neighbor interactions from the 3 × 3 × 3 interaction template need be considered. This number 13 is the same as the number of chaining cells which have to be considered in Hockney's PPPM data structure to find all objects within the cutoff radius R_c. Figure 3.5 shows a schematic rendition of these different interaction templates. Only the half of the template with index offset larger than zero has to be considered since all interactions with objects having a lower storage address index will have

been calculated previously. As shown, shells of interaction can be defined which will correspond approximately to neighbors at different physical distances. The 16 neighboring nodes indicated with grey squares form the closest shell. The 30 triangle nodes are a bit further away, on average, and the 16 circle nodes form the furthest shell of the nearest neighbors template.

SECTION IV. ADDITIONAL ASPECTS OF MONOTONIC LOGICAL GRIDS

IV.A Possibilities for Further Optimization

By a fully vectorizeable process of exchanging or "swapping" objects between adjacent logical cells every few timesteps, the nearest-neighbor MLG ordering is kept intact even though the objects move from cell to cell. Thus the method can be applied to gas, solid, and liquid systems using the same logical structure for problems of interesting size, i.e. 1,000 - 10,000 particles. The MLG algorithms forego a regular grid in space with a variable number of objects in each cell for an irregular spatial grid which has exactly one object per cell by construction. This logical simplification, brought about by the MLG mapping, permits extensive optimization under current and planned supercomputer architectures (e.g. Fox and Otto, 1984) without sacrificing the generality needed to make it useful.

Optimization of nearest-neighbor algorithms for particle dynamics is both machine and problem dependent. Vectorization techniques to achieve very high rates of computation require that all logical and arithmetic operations be performed on organized arrays of independent data. Distributed processing approaches to massive parallelism rely on a number of self-controlled processing centers operating asynchronously, but according to fixed rules of cooperation, on an evolving data base. To take advantage of both approaches simultaneously requires being able to define a number of vectorizeable segments of the problem which can be calculated independently. Furthermore, the vectors must be long enough to be computationally efficient but short enough that the memory needed in each asynchronous processing

center is not prohibitively expensive. The MLG algorithms presented in the previous section can be partitioned for multi-tasking across a number of independent processors.

To maximize the length of vectors within each partition when the typical MLG dimension, $NX = NY = NZ = N^{1/3}$, is only about 20 (8,000 objects) requires treating a substantial fraction of a plane as a single vector. In the $8 \times 8 \times 8$ test problem, vectors of length 64 can be used throughout except for the X-direction monotonicity tests where vectors half as long would result. This is accomplished by collapsing several indices into one index and by paying careful attention to the boundary conditions.

Optimum-computational efficiency results when the last few swapping operations are performed only for the grid points which might have become non-monotone due to adjacent swaps taking place during the previous iteration. Reductions of up to a factor of two in computer work to maintain monotonicity might be obtained by reducing the number of inactive vector swap attempts greatly at the cost of considerably increased program complexity. The scalar program to perform the few remaining swaps and keep track of which few nodes might have had their monotonicity conditions affected by the previous swaps is complicated. To date, convergence of the number of vector swapping iterations required has been so fast that this extra work has not been indicated. In the future it may be worth the effort for production calculations.

The same kind of gain can be obtained by trimming the near neighbors template defining which logical neighbors are likely enough to be close spatial neighbors that they should be included in the vector interaction calculations automatically. When a scalar "clean-up" portion is added to

the vector force summing algorithm, the number of logically neighboring nodes which are always considered can be reduced significantly below that required to ensure no close uncounters. Figure 3.5 shows three shells of logical interactions in the nearest neighbors template, each succeeding shell taking neighbors which are logically, and usually physically, farther away.

By keeping track of maximum X, Y, and Z displacements along each row, column, and plane in the nearest neighbors template, a smaller template can be used with assurance that there will be no close "uncounters". After performing the interaction analysis on the objects within the vector shell, the boundaries of the shell can be checked to ensure that objects on the logical boundary are far enough away spatially. These checks over all the objects in the grid can also be vectorized. If the template has been pared sufficiently to ensure a worthwhile reduction in the number of objects that have to be considered most of the time, there will generally be a fraction for which one or more of the neighbors in the vector shell were not far enough away to ensure that the next neighbor, which is outside the vector shell, can safely be neglected.

For the few objects which may have spatially close neighbors which are removed more than two or three locations logically, a scalar calculation can be performed. It would probably save a factor of two or so in overall computational cost to reduce the vector shells of the nearest neighbors template until the scalar cost competes with the significantly reduced vector cost. The expense of keeping track of X_{max}, Y_{max}, and Z_{max} are minimal, three vector operations per object per near neighbor interaction. The test to determine which few objects require extra (scalar) work is even

/5-

cheaper, a few vector operations per object. Once an object has been found to require extra work, the scalar search can be extended to whatever logical distance is necessary to ensure that physically nearby objects do not go uncounted.

A 5 × 5 × 5 cubical nearest neighbors template has 62 interactions which will be considered for each object. From empirical evidence to date this is adequate provided the critical radius of consideration is somewhat less than the average separation, here taken to be the original object spacing. Relatively few close uncounters can occur because neglected objects are logically at least 3 and generally 4, 5, or 6 nodes away.

Holes can be added to the MLG, locations which move or stay fixed in space but which don't contain an object. Any object neighboring one of these holes has one fewer real object in its interaction template because of the hole but this obvious disadvantage is balanced by the fact that hole locations can be updated any way necessary to improve the locality and structure of the MLG. By adding or shifting holes about judiciously it may be possible to avoid highly distorted MLGs. The holes would be subject to swapping with objects just as if they were objects but their equations of motion can be different and their interactions with real objects zero.

Figure 4.1 was computed using the 512 particle model with point non-interacting particles and the complete 5 × 5 × 5 interaction template shown in Figure 3.5. The volume around each particle was divided into shells of thickness 1 Angstrom and the number of particles in each radial shell was counted for particles logically outside the 5 × 5 × 5 template to determine how often "close uncounters" occur. A close uncounter occurs when an object gets close physically to another object without coming within the MLG

nearest neighbors template and thus ensuring that the interaction is "counted" in the vector sum. Concentrating first on the common features of the two physically identical calculations shown in the figure, we see that probability of an uncounted particle penetrating the interaction volume drops off very rapidly as the distance becomes small and hence the physical interaction would be important. It is one hundred times less likely to find an undetected particle coming within 10 Angstroms than to find one coming within 20 Angstroms. It is another one hundred times less likely to find one coming with 5 Angstroms and no close uncounters were ever found less than about 3 Angstroms.

The problem was repeated with a ninth plane of 64 locations added to the calculation for holes. The holes were given the average location of their six nearest logical neighbors as a propagation law and rapidly mixed throughout the volume with the particles. As can be seen, the probability of unrecorded close encounters is essentially unchanged. In this test, at least, holes don't seem to help much. This is probably because the propagation law used took no detailed account of local grid irregularities signalling a possible "close uncounter" problem. Clearly research is needed to refine this generalization to the point where it appreciably optimizes the MLG representation.

IV.B Fluid Dynamic Applications of the MLG

A few words about the application of the MLG to Lagrangian fluid dynamics is appropriate here. Each node of the grid can be identified with a fluid or vortex element. The advantage is in having a regular grid available to solve the physical evolution equations. Elliptic equations,

for example, become amenable to highly efficient, vectorized multigrid methods (DeVore, 1984) on regular $N_x \times N_y \times N_z$ grids even though the fluid elements themselves move randomly. Two-dimensional and four-dimensional problems can be handled just as easily by the same methods.

Work is needed telling how to evaluate spatial derivatives accurately on the distorted MLG. When points are far apart spatially, the fluid cannot be as accurately represented as when they are close. To keep the resolution more nearly uniform than the specific fluid flow may be capable of, it can become necessary to remove nodes where they are crowded and to inject them elsewhere to better resolve some regions. To do this in the MLG involves finding a fluid element which can be merged with a larger one nearby in a manner which conserves mass, momentum, and energy. This frees up a location which can be "shifted" to the correct row, column, and plane to improve a deteriorating local resolution. The process in 2D is quite analogous to the operations needed to order sliding tiles numbered 1 to 15 in the 4 × 4 spaces of a popular child's game. The hole is shifted to the place where it is most needed.

After a shift operation, the same as an ordered series of swaps, it is likely that local swapping may be necessary to reestablish monotonicity. Shifting a line in one direction may well trigger swaps in the other two directions, so the real cost of adding and subtracting Lagrargian nodes locally to control resolution has to be measured (or estimated) for each configuration being considered. Certainly the specific additions, deletions and shifts will have a significant scalar component of computation. It does not make sense to execute a vector swapping iteration over the whole grid unless a significant fraction of the objects are being swapped. However, as

long as at least 5% of the objects are being moved, the overall vector swapping iteration probably pays.

For comparison consider another free Lagrangian approach, the Lagrangian Triangular (Tetrahedronal) Grid (Crowley, 1972; Boris and Fritts, 1975; Fritts and Boris, 1979). In this approach the logical grid structure varies in time as the nodes move. The number of nearest neighbors can vary from node to node and the number and identity of these neighbors can vary at a given node as the Lagrangian configurations change. This extra freedom, not allowed in the MLG, is used to maintain a local grid structure optimized to guarantee diagonal dominance of the simplest conservative finite-difference elliptic operator. The price is the loss of local order in the grid and hence no vectorization.

Generalizing this Lagrangian Triangular Grid (LTG) to 3D is straightforward but operationally very complicated. The grid is composed of adaptively restructuring arrangements of tetrahedra in this case. The local grid structure can still be changed as needed to maintain diagonal dominance of the elliptic operator and resolution can be increased or decreased locally as needed. Scalar linked lists become a necessary evil to keep track of nearest neighbors though the resulting algorithms are still of order N.

Clearly the local spatial structure of the MLG is not as "good" as in the generally structured LTG but the global structure compensates for this. The monotonicity conditions specify a meaningful and useful relationship between spatial derivatives and grid differences. As a result, fluid flows with long range correlations, unlike the random particle motions used in earlier tests, may lead to an additional computational expense at specific

times. In the smooth flow of large rotating and translating vortices, an initially rectilinear grid might survive many timesteps before any of the local monotonicity constraints are violated. Nearby points would move in almost the same way. Once the fluid rotates far enough, however, monotonicity violations would have to occur. Because of the long range correlation of the motions, a number of swapping iterations may be necessary to reset the MLG.

In a turbulent flow with coherent flow structures at several spatial scales, we can expect intermittant bursts of swapping activity from different scales at different temporal frequencies. Small coherent structures in the flow require fewer swapping iterations to restore monotonicity but will require them more often than large structures when the rotation rates are higher. Though the integrated number of these swapping iterations is large, it is unlikely to be larger than the number of swaps required for random object motion with the same typical distances traversed.

The strong possibility exists that statistics on the number and frequency spectrum of grid swaps may prove to be a very good diagnostic of the progress of turbulent mixing and flow interpenetration. Each swap can be viewed as a single quantum of geometrical upset. The object swapping process is necessarily intermittant and thus presents the possibility of easy integral measures of local intermittancy and turbulence spectrum. Studying the object swapping time series might provide a very direct way of diagnosing such fluid simulations for the onset and character of chaos.

IV.C Other Applications of a Monotonic Logical Grid

The MLG suggests itself for use in multiphase fluid problems. Each grid node could be used to represent a droplet in a spray or a grain of sand in a sandstorm. Droplets could have varying sizes which increase or decrease in time due to local surface effects like condensation, evaporation, or abrasion and all the droplets would not have to be simulated. The accumulation of temporal averages over times and distances short compared to changes in the background flow means that only a small fraction of all the droplets or particles would have to be followed in the MLG to get a good estimate of the interactions of the whole distribution of particles with the background gas.

Collisions of two drops could occasion merging or fragmentation. The MLG can accomodate either by shift operations which transport nodes from where they are no longer needed to new sites where fresh droplets are formed or enter the system. The gaseous background could be represented on a Eulerian mesh to facilitate swapping of mass, momentum and energy back and forth between particles and gas. The volume overlap of MLG cells with cells of the Eulerian grid can be used to circumvent a major complication of Monte Carlo methods, choosing the values of continuum functions at places where there are no particles or Lagrangian nodes. The MLG cells provide a natural way to interpolate back and forth between the two representations.

When insolation of dust or droplet clouds is important, the MLG provides a simple way to assess the radiation opacity along any particular direction. The grid axes can be chosen in a given direction and the swapping algorithm used to resort the points along that direction. There

are a number of line-of-sight obscuration problems where this flexibility will be useful.

The Monotonic Logical Grid (MLG) has been given only a geometric context so far. In the applications and examples above, the moving objects are being arranged relative to each other in 3D Cartesian space. Other more abstract applications suggest themselves. The MLG can just as easily represent multidimensional phase spaces for Boltzmann and Vlasov Equations. The grid may also be useful for some classes of problems involving more abstract data organization. The MLG is constructed using a set of monotonicity operators, one for each dimension. These are really relational or comparison operators that return a logical "yes", "maybe", or "no" depending on the results of a generalized comparison between two logical entities. The operators can be black boxes implementing complex, possibly subjective comparison algorithms which need have nothing to do with geometry or physical nearness. The meaning behind the relational operators can be almost anything; for example, "degree of difficulty", "technical merit" and "artistic merit" may be separate awards in the rating of an athletic contest.

V. SUMMARY AND CONCLUSIONS

This report introduced a simple, vectorized algorithm to determine nearest neighbors whose cost scales as the number N of independent objects or locations. This is accomplished by defining a Monotonic Logical Grid (MLG) for storing the object data dynamically so that objects which are adjacent in real space are automatically close neighbors in the logical grid as well. As a simple geometric test problem, a regular $8 \times 8 \times 8$ 3D grid was used to store the position and velocity components of 512 randomly located particles in a cubical domain. For this idealized system the points were given random velocities and the MLG was evolved for many transits of the system by the faster particles. Statistics on near encounters of logically far away points and on the number of restructuring operations required were presented.

It was found that the reconnections of the dynamically changing MLG can generally be computed locally in a very few vectorized iterations without using inefficient gather or scatter operations. Almost all of the grid restructuring occasioned by particles passing each other occurs in the first two or three vectorized iterations. Further optimization is possible by changing to local scalar swapping after a few iterations. It is also found that almost all the spatially closest nodes are nearby in the MLG as well. Two or three logical grid locations effectively defines the spatial near neighborhood except for a vanishing small number of cases which can be detected and corrected inexpensively.

The MLG differs from previous nearest-neighbor algorithms. It effectively removes the constraint of having to associate a cell of the

logical grid with a fixed region of real space, but introduces the constraint of only one particle per computational cell. When many of the objects cluster somewhere, a corresponding fraction of the storage locations in the MLG are automatically associated with that region. This means that substantial variations in object density are adaptively gridded by the MLG and large regions of space, as well as computer memory, are not occupied by empty cells.

This algorithm gives regular global orderings of the object data and so allows efficient contiguous vector operations which are longer than the relatively small number of neighbors considered for each object but can be much shorter than the total number N. The algorithm will execute efficiently in small array processors and permits direct partitioning to take advantage of massive asynchronous parallelism in VLSI/VHSIC-based distributed processing systems. The cost to execute the simplest version of the model is one hour on a DEC VAX 11/780 for one thousand particles for 1000 timesteps when a simple force law for the 124 nearest neighbors is used. With a commercially available Distributed Processing System, 5000 interacting objects could be monitored and the data base updated and then restructured thousands of times in about 15 minutes, fast enough for realtime applications.

A number of potential applications were discussed briefly. Obviously other uses will suggest themselves as the good properties and restrictions of the interesting multivalued geometric MLG mapping between real space and relative (logical or computer storage) space becomes better understood. These problem-independent properties will be necessary to the successful

application of the representation to practical problems with other mathematical, logical or physical constraints. Swapping and shifting operations and holes were introduced to allow efficient local and global grid readjustments. Practical experience with the MLG is still small, however, so major pathologies may yet be uncovered in some applications.

Many MLG configurations may be possible for the same physical node arrangements and simple examples suggest that the best configurations are much better than the worst. Thus efficient methods of optimizing local and global structure within the monotonicity constraints will eventually be imperative. Additional work is needed on the following questions:

1. What is the mathematical nature of the simple representations for spatial derivative operators and integral conservation operators and how can they be optimized computationally?
2. Is there an algorithm to optimize the grid structure using holes and/or adaptively varied local modifications of the monotonicity functions?
3. What is the cost of not reaching monotonicity every cycle?
4. What is the geometric or information theoretic meaning behind the ambiguity of possible representations, i.e. what kind of an uncertainty principle does this represent?

16

Acknowledgements

I would like to acknowledge years of informative and rewarding discussions with Dr. Martin Fritts on topics ranging over all aspects of this subject. His diligent and creative efforts on the LTG approach have provided an information base for the development of the MLG algorithm. I would also like to thank Dr. Sam Lambrakos for providing the information in Figures 3.3 and 3.4. This work was supported by the Office of Naval Research projects in Large-Scale Scientific Computing (44-1909, RR014-03-05), Computational Hydrodynamics (44-0573, RR01403-02) and Molecular Dynamics (44-1950-0-5, 61153N), and by the Naval Research Laboratory.

REFERENCES

1. R.W. Hockney and J.W. Eastwood, "Computer Simulation Using Particles", Chapter 8, pp267-304 (McGraw-Hill Inc, New York, 1981).

2. W.F. Gunsteren, H.J.C. Berendsen, F. Colonna, D. Perahia, J.P. Hollenberg, and D. Lellouch, "On Searching Neighbors in Computer Simulations of Macromolecular Systems", Journal of Computational Chemistry, Vol. 5, No. 3, pp272-279 (1984).

3. G.C. Fox and S.W. Otto, "Algorithms for Concurrent Processors", Physics Today, pp50-59, May 1984.

4. L.M. Adams and H.F. Jordan, "Is SOR Color-Blind?", ICASE Report No. 84-14, NASA Langley Research Center, May 1984.

REFERENCES CONTINUED

5. J.P. Boris, M. Fritts, and K.L. Hain, "Free Surface Hydrodynamics Using a Lagrangian Triangular Mesh"; Proceedings of the First International Conference on Numerical Ship Hydrodynamics, Gaithersburg, MD, October 20-22, 1975.

6. M.J. Fritts and J.P. Boris, "The Lagrangian Solution of Transient Problems in Hydrodynamics Using a Triangular Mesh", J. Comp. Phys., Vol 31, No 2, p173, May 1979.

7. W.P. Crowley, "FLAG: A Free Lagrange Method for Numerically Simulating Hydrodynamic Flows in Two Dimensions", Proceedings of the Second International Conference on Numerical Methods in Fluid Dynamics, (Springer-Verlag, New York, 1971).

8. C.R. Devore, "Vectorization and Implementation of an Efficient Multigrid Algorithm for the Solution of Elliptic Partial Differential Equations", NRL Memorandum Report, October, 1984.

16

FIGURE CAPTIONS

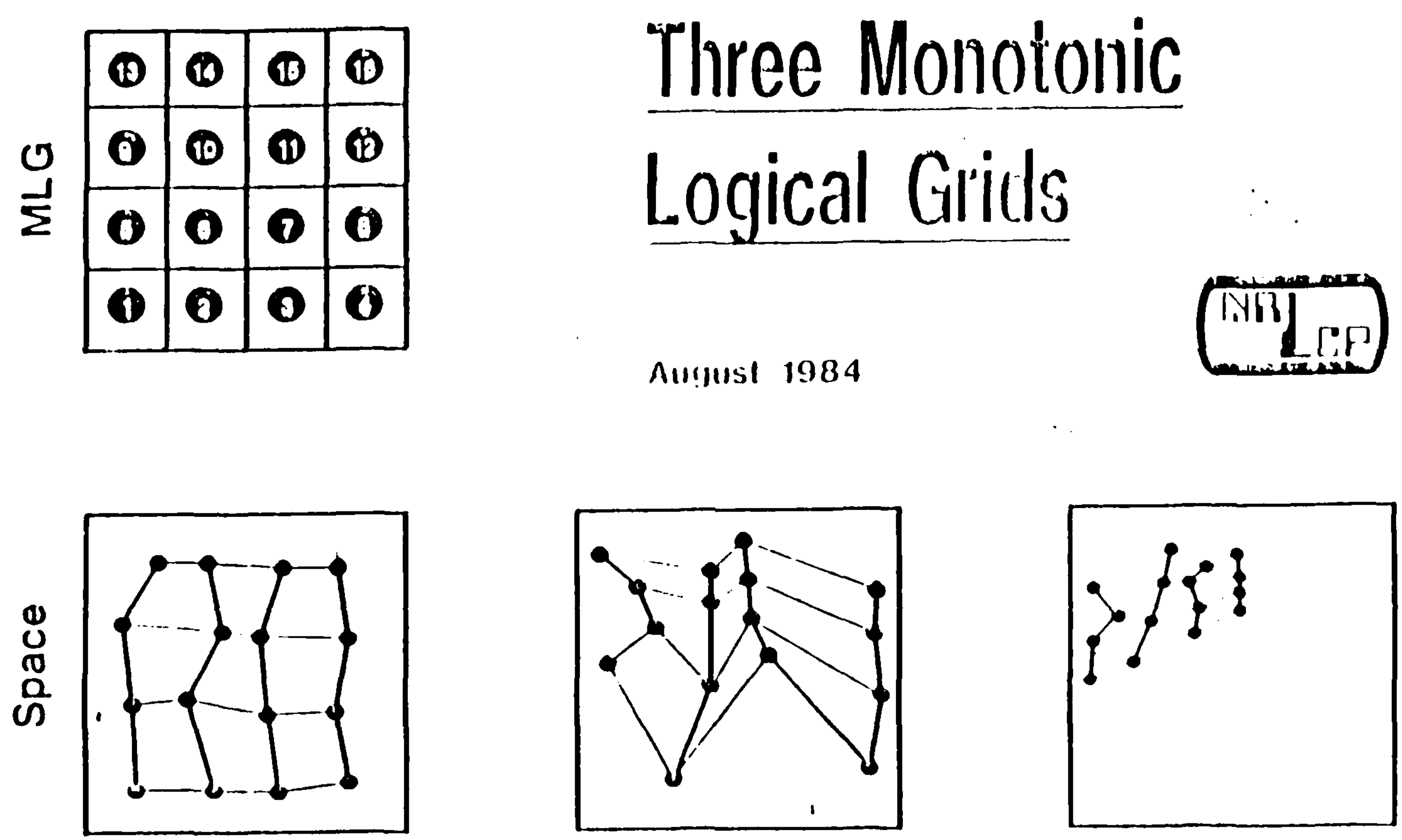
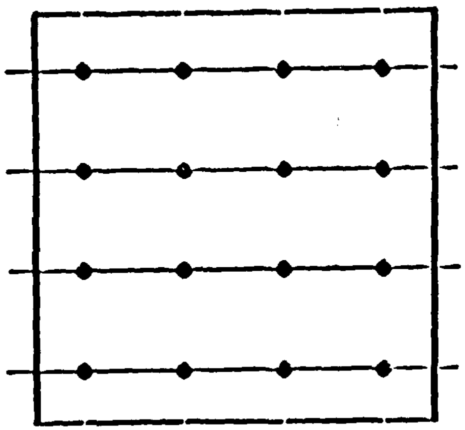


Figure 3.1 Three Different Spatial Configurations and the Corresponding Monotonic Logical Grid



Three Monotonic Logical Grids from Identical Data

August 1984

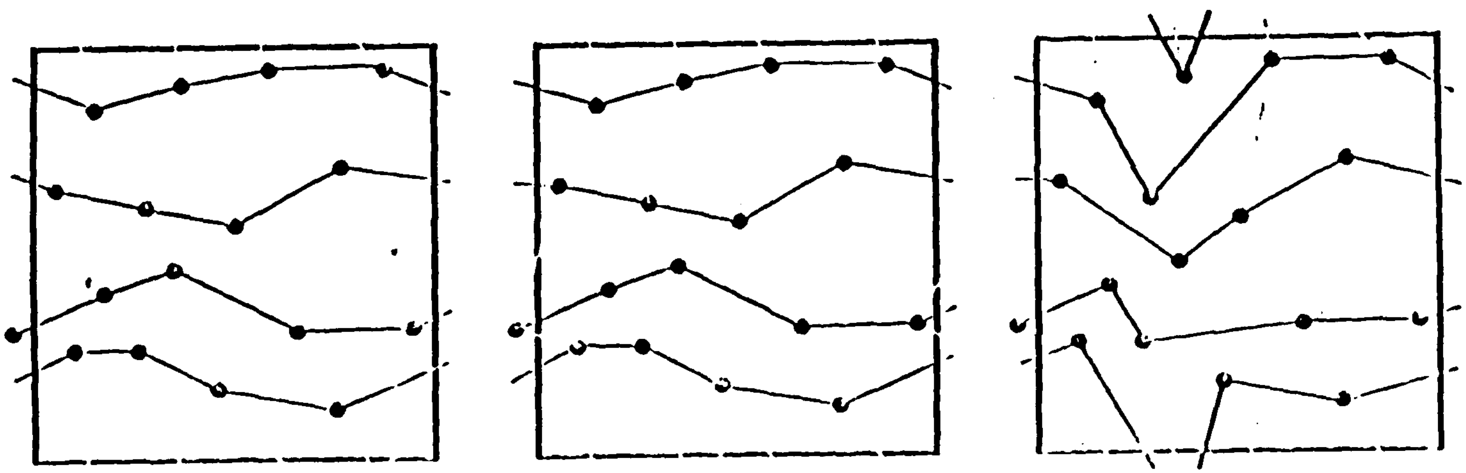


Figure 3.2

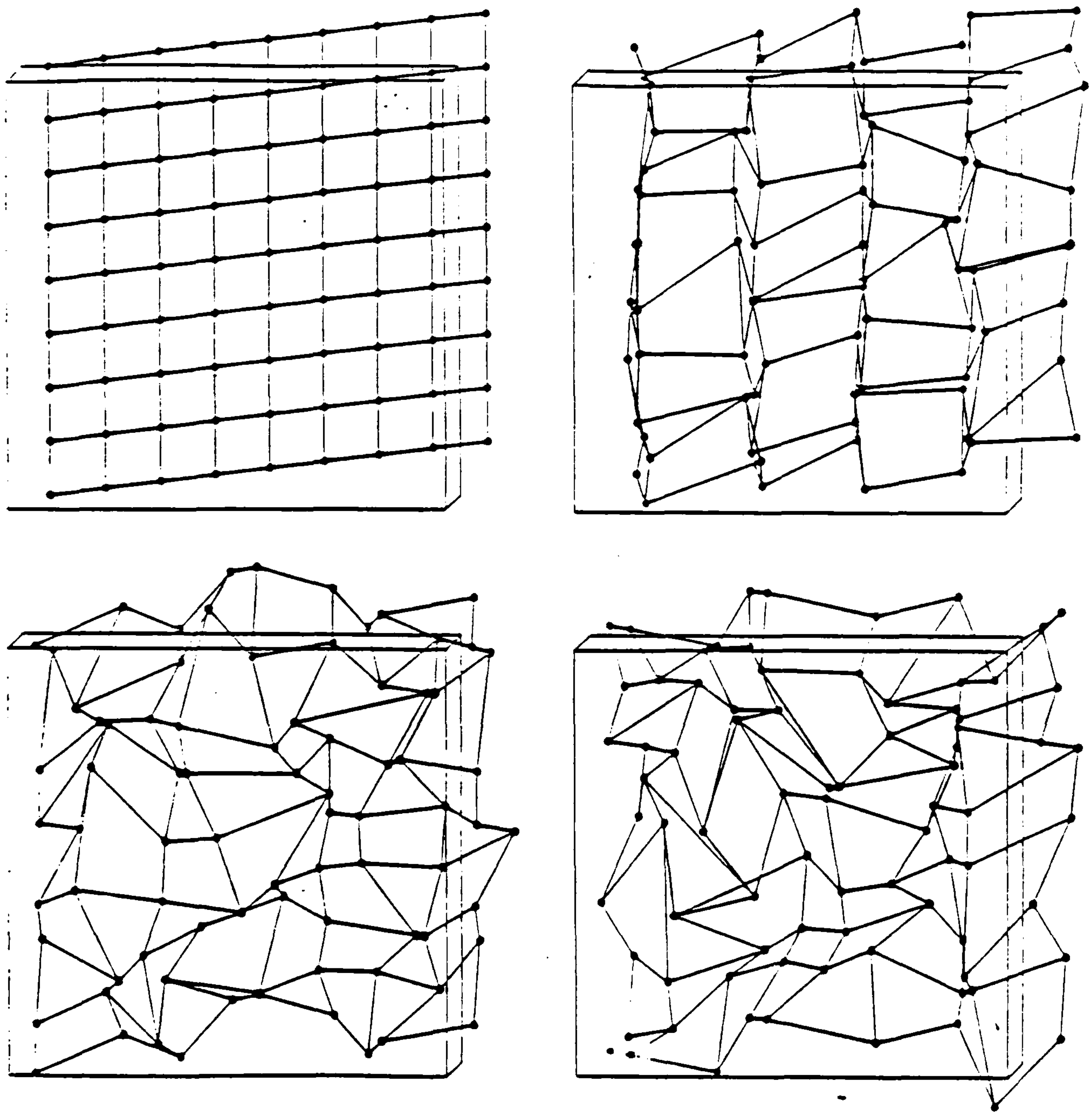


Figure 3.3 The MLG for One Plane of the 512 Particle Test Problem

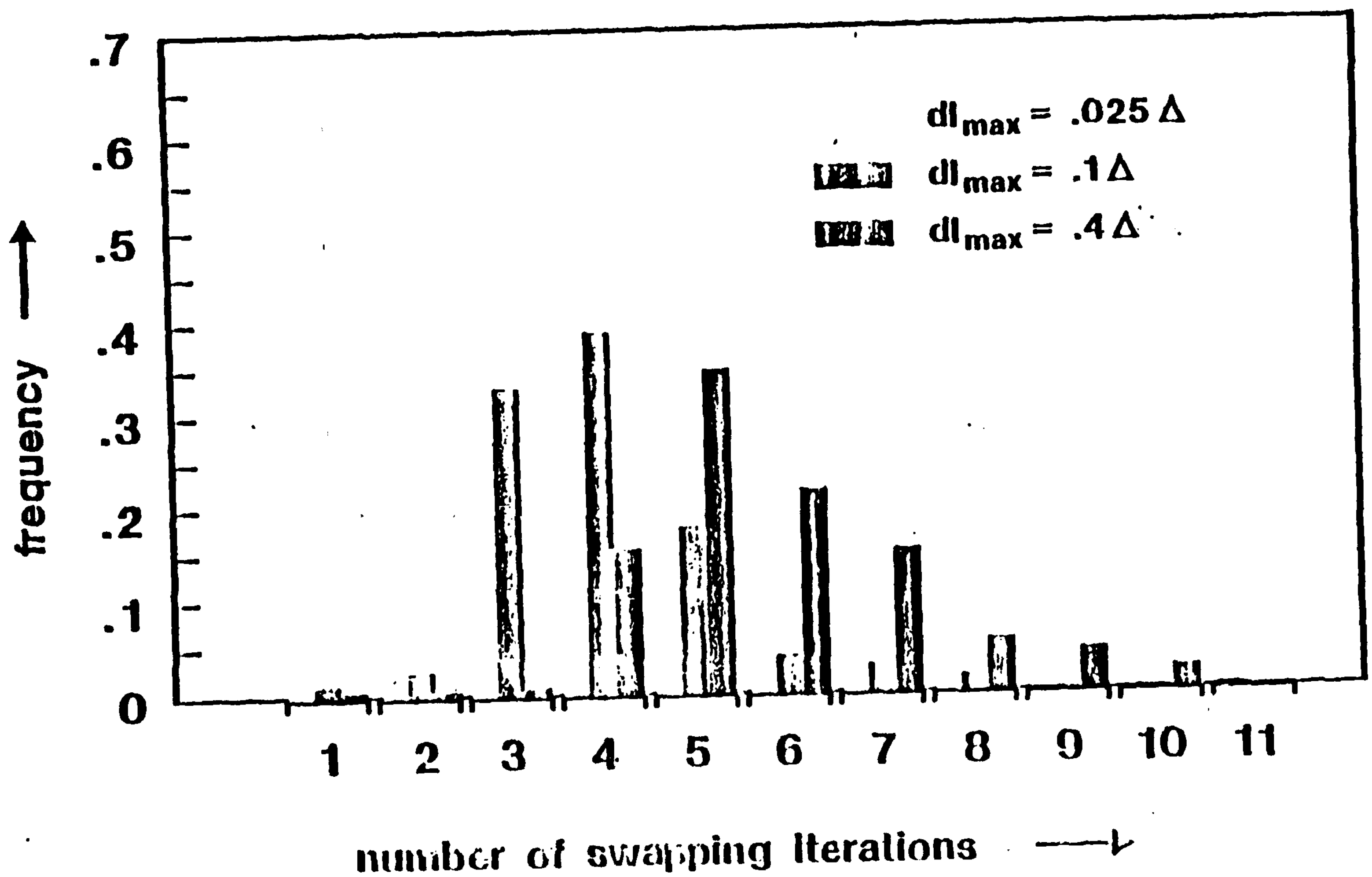
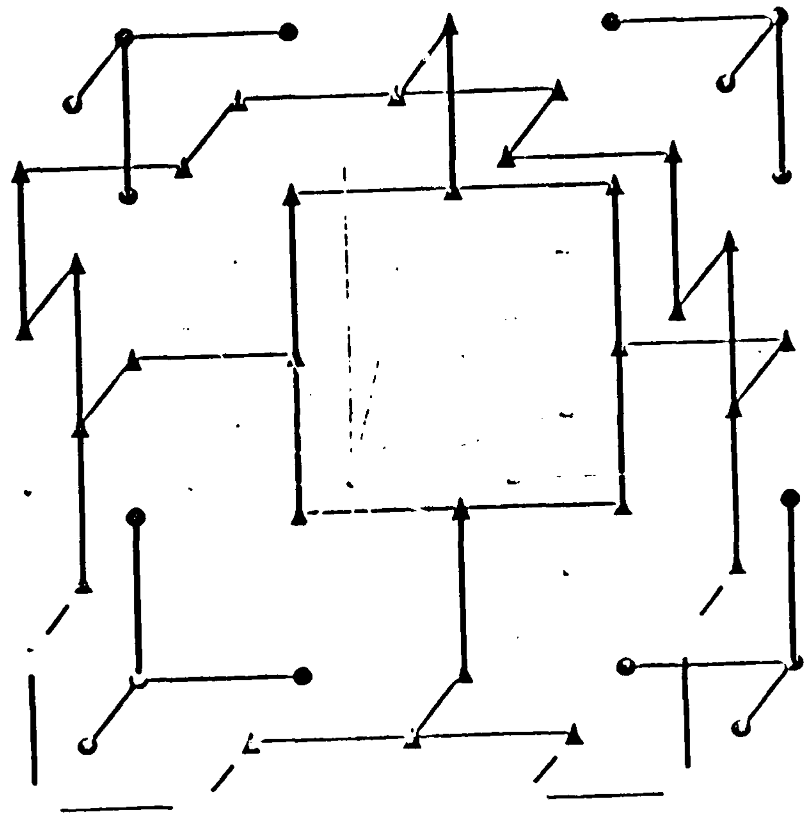


Figure 3.4 Frequency of Requiring N Swapping Iterations to Construct a Monotonic Logical Field

LOGICAL DISPLACEMENTS OF NEAREST NEIGHBORS

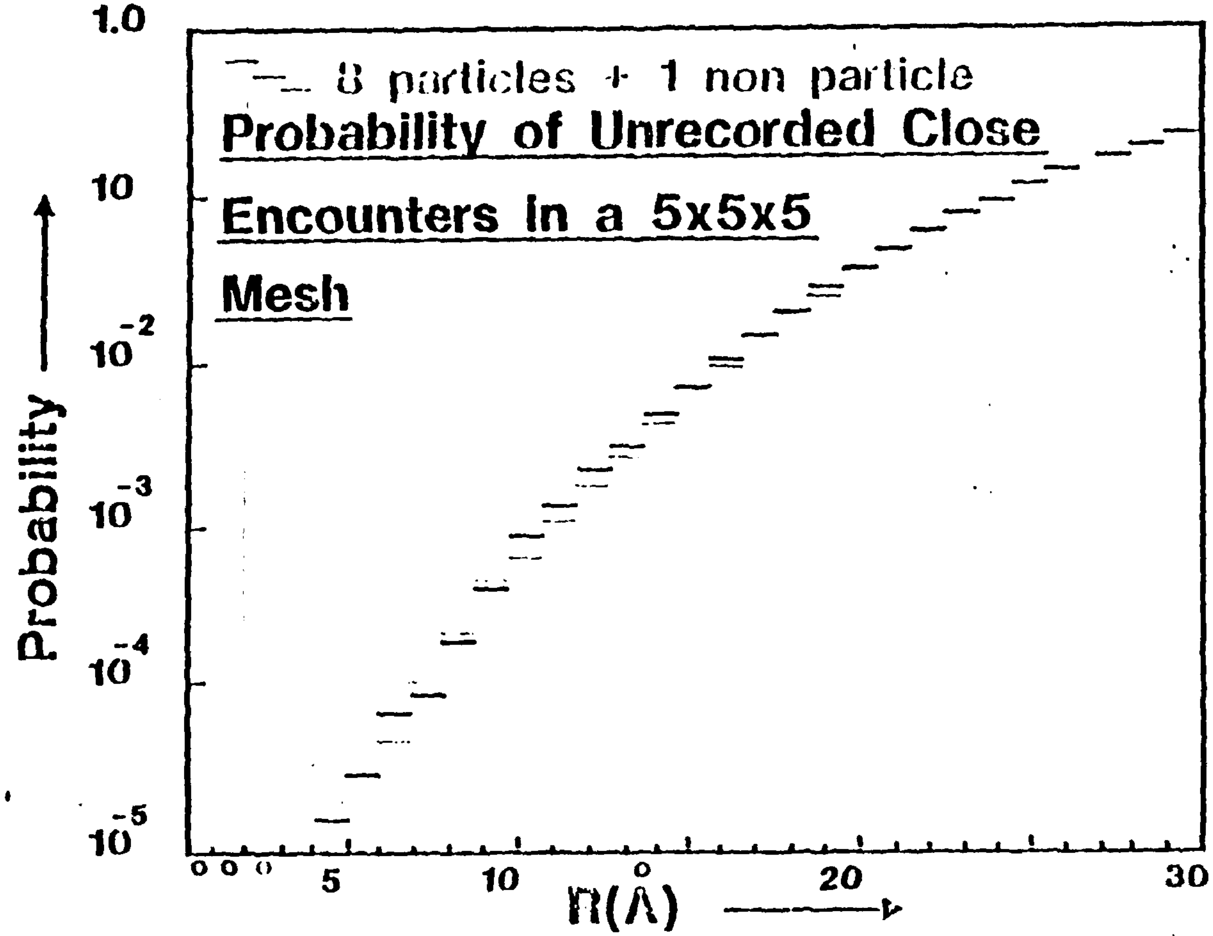


NRL LCP

August 1984

Figure 3.5

175



August 1984

Figure 4.1

LJI-R-85-349

FOUR PAPERS ON ROBOTICS:

I. WHITE PAPER FOR THE DEVELOPMENT OF THE "SUPER ROBOT"

II. NEXT GENERATION OF TECHNOLOGY FOR ROBOTICS

III. SUMMARY REPORT ON PIPELINED COMPUTATION OF DYNAMIC MODELING MATRICES FOR SERIAL ROBOTIC MANIPULATORS

IV. AN ASSESSMENT OF THE DEVELOPMENT AND APPLICATION POTENTIAL FOR ROBOTS TO SUPPORT SPACE STATION OPERATIONS

BY

DELBERT TESAR*

DECEMBER 1985

LA JOLLA INSTITUTE
P. O. BOX 1434
LA JOLLA, CA 92038
(619) 587-6000

THIS RESEARCH WAS SPONSORED BY THE
DEFENSE ADVANCED RESEARCH PROJECTS AGENCY
UNDER ARPA ORDER NO.: 3710
CONTRACT NO.: MDA903-85-C-0187

*Carol Cockrell Curran Chair in Engineering, Mechanical Engineering Department, College of Engineering, The University of Texas at Austin

174

I. WHITE PAPER FOR THE DEVELOPMENT OF THE "SUPER ROBOT"

BY

DELBERT TESAR

Concept

The super robot represents the full integration of the most advanced computer technology (the super computer) with the most general mechanical architecture (serial, parallel, modular, etc.) to demonstrate an electronically rigid system (similar to our latest fly-by-wire aircraft) capable of rejecting process disturbances in real time while producing high value-added products on demand. Today, high value-added operations are achieved primarily through the use of expensive, specialized and dedicated machines such as N.C. machines, automatic screw machines, wire bonding equipment for micro-circuits, etc., where the robot performs the low-valued function of handling of parts between these dedicated machines. By contrast, the super robot would be a fully integrated and self-contained generic machine system capable of performing a wide spectrum of precision light machining operations completely programmable by the designer of the product (shoes, clothes, appliances, etc.) and fully responsive to the individual demands of the marketplace. This vision of robotics by Issac Asimov is the heart of the factory of the future, yet it not only does not exist, technical resources to make it possible are either in short supply or have not been concentrated in a sufficient critical mass of expertise to make it happen.

Beyond the factory of the future there are applications of robotics to functions which involve hazards to humans such as space operations, operations on the ocean floor, ammunition handling under chemical or biological attack, processing of dangerous materials such as galium arsenide for advanced microcircuit technology, nuclear reactor maintenance, etc. In addition, special applications of real value to society, such as micro-surgery, have yet to be dealt with even in the research environment. The concept of the super robot being proposed here would lay the foundation to demonstrate a science of intelligent machines sufficiently general to treat all of these diverse and rewarding applications.

Justification

Much of the mechanical design philosophy in the United States derives from a period during which farm machinery, power plants, construction machinery, automobiles, airplanes, jet engines, etc. were brought to a high level of development. Much of this design is performed in terms of compartmentalized rules (the basis of an art and the opposite of a science) which are based on negative criteria (noise, wear, fatigue, instability, vibrations, mean time between failures, etc.). On the other hand, the factory of the future demands the use of operational criteria associated with the quality of the product of the machine which implies precision (rarely dealt with as a first priority in the academic world). The positive criteria of precision involves the control of the output of the machine to specified tolerances regardless of the disturbances generated by the operation. To date not a single robot operates in terms of a real time dynamic model based on an accurate description of its system parameters in order to reject disturbances (i.e., the concept of closed loop operation found in fly-by-wire aircraft). Furthermore, the negative design criteria of failure in the operation of large machine structures of the past (textile machinery, battlefield materiel, etc.) have little to offer for the design of precision micro-processing equipment of the scale suitable to micro-surgery or micro-circuits. Hence, relative to the level of technical integration required to meet future needs, no balanced science of intelligent machines is being developed.

Today, the drive to establish the factory of the future has led to vigorous development activity associated with CAD/CAM. Unfortunately, almost all of this activity is centered on the use of a collection of dedicated machines each capable of a limited number of distinct critical precision functions which must be sequenced to create the finished product. On the other hand, the fully integrated self-contained

intelligent machine which is capable of producing broad classes of quality products fully responsive to the individual consumer does not exist in any form. In fact, the use of 15,000 robots in the U.S. at this time implies a penetration into the manufacturing workplace of not more than 1 in 1000 showing that robot implementation is far below the level needed to have real impact. This low level of penetration may be due partially to the fact that each of our major firms (IBM, GE, GM, Westinghouse) made one robot and then decided to purchase robots from outside vendors or to buy subsidiaries either in Japan, Europe, or in the U.S. By contrast, in Japan, each of the major manufactur .g firms (Hitachi, Mitsubishi, Fujitsu, etc.) make their own robots. The content.on here is that U.S. firms do not have the necessary in- house, balanced technical manpower to remain competitive in this leading edge technology and are leaving it to their economic competitors. This lack of response to the threat of the trade deficit, exceeding $100 billion in value-added products, is at the heart of the present proposal. The goal is to employ existing component technologies (the super computer, computer vision, digital control theory), enhance emerging technologies (expert systems, artificial intelligence, metrology, mechanical architecture, computer architecture, CAD/CAM), and fully integrate them by means of a balanced science for intelligent machines. The super robot would be the most aggressive demonstration of this objective.

Background

The simplest robotic architecture is a 6 degree-of-freedom (DOF) serial system (one link, one joint, one link, etc.). To date two basic geometries have emerged. One is a structure similar to a coordinate axis (X-Y-Z) machine and the other is similar to a human arm. These simple structures are used because they represent very few design parameters and are designed primarily by intuitive means. The general 6-DOF serial robot system is described by 18 geometric, 36 mass, 36 deformation, and 18 control parameters (a total of 108) and represents a design complexity far beyond the means of existing expertise in industry. Beyond the serial structure, there are parallel structures (walking machines with 4 or more legs), redundant structures having excess prime mover inputs or excess degrees-of-freedom, modular structures to form systems from building blocks the way we now create computer systems, etc. What this means is that the design techniques for most future robotic systems do not yet exist and can only be developed by a very aggressive research program.

Similarly, no industrial robot operates in terms of a real time dynamic model description to close the loop relative to the process it is performing which may generate significant disturbances in the system. This means that precision light machining operations such as drilling, routing, milling, etc., cannot be performed by reasonably sized generic robots to the level of precision required. Disturbances due to forces equivalent to the specified load capacity of these robots can easily cause a deflection 20 times as great as the error represented by its repeatability (i.e., a 20 to 1 robot). The goal must be to measure these disturbances and to compensate for the resulting deformations (in order to maintain the desired level of precision) by means of a complete dynamic model evaluated in less than 10 milli-sec. (real time) by using the most modern computational hardware and software. This class of control would be equivalent to feed forward compensation (a technique now found in the very best Japanese Hi-Fi equipment) and is what is meant by an electronically rigid robot system.

Over the past several decades, the electrical research community has made major strides forward in its technical depth especially enhanced by strong "pulls" from the civil and defense sectors. By contrast, mechanical technology has not kept pace such that it is now perceived as a weak partner. Unfortunately, the mission objective of intelligent machines will require a marriage of these technologies as equals. Hence, in order to satisfy the super robot development mission, it will be essential to create a fully integrated science of intelligent machines based on a balanced development of all required electrical and mechanical component technologies.

Technical Objective

The concept of the super robot is the full implementation of a cohesive analytical description of generalized mechanical architecture with a major emphasis on the use of the super computer to benchmark the complete controlling equations for deformation, dynamics, adaptive control, and feedforward compensation for the effects of external or process generated disturbances in real time operation. The objective is to obtain results which are able to describe the operation of any general robotic structure thus allowing for specialization to a given device suited to a unique application. This top-down approach (similar to the approach used to validate the development of the supercomputer itself) is completely missing in the development of robotics to date. Literally hundreds of design parameters are involved yielding potentially billions of possible systems. The optimal design and, therefore operation, of these systems is essentially unreachable with present small scale bottom up technologies. That is why most industrial robots look either like an X-Y-Z measuring machine or a human arm both of which are several orders of magnitude simpler than the general mechanical architecture. Consequently, the super-robot effort is intended to integrate all the previous analytical research of the team (and that of others) into a fully operational simulation test-bed for both design and operation of completely general serial, parallel, modular, or mixed robotic structures.

The following is a partial listing of applications which would become feasible or would be dramatically accelerated by development of the super-robot technology:

1. Nuclear reactor maintenance
2. Precision light machining
3. Micro-manipulation at very small scales
4. Micro-surgery
5. Ocean floor operations
6. Space station operations
7. Battlefield operations
8. 50G centrifuge robot
9. Rapid runway repair even under attack
10. Remanufacture of military hardware such as jet engines, airframes, etc.
11. Walking machines and cooperating robots
12. Human augmentation for the handicapped

Research Program

The research program will concentrate on the use of the super computer to dramatically accelerate the development of a science of intelligent machines because of its superior computational capacity to treat the full parametric description of a much more general class of robot structures. For example, the massive computational resources of the supercomputer makes it possible for the researcher to think much more openly and freely of generic top down design and control strategies which should lead to a maximum level of productivity of new ideas and technology evaluated by complete simulations. This increased computational capacity will mean that the following can be addressed:

1. Metrology of Robots - A semi-automatic means of identifying all significant parameters in an existing robot.

2. Optimal Design - Initial success in the use of optimization techniques to the multi-parameter multi-criteria problem associated with robotics has led to improved distribution of actuator parameters. This computationally intensive effort must be expanded.

178

3. Robot Architecture - Future robots will be composed of easily scaled structural modules (shoulders, wrists, micro-manipulator, mixed large and small control structures, etc.) to provide finite packages of proven technology to be rapidly assembled into generic intelligent machines.

4. Computer Architecture - The top down approach made feasible by the supercomputer will make it possible to develop specialized computer hardware and software modules (arithmetics, array processors, etc.) uniquely suited to intelligent machines.

5. Operational Software - Symbolic programming can now be applied to the complex analytical formulations required to completely describe the dynamic state of a robot and to form the basis for a generic operational language capable of off-line programming and disturbance rejection.

6. Adaptive Control - This represents the real time adjustment of the control parameters to best enhance the controllability of the fully nonlinear nature of robot structures.

7. Walking Machines - These parallel machines are essentially multiple robots cooperating to perform a finite number of output functions with excess (a factor of 3 or 4) of prime mover input commands. This becomes a real time resource allocation problem (artificial intelligence) of the most complex nature.

8. Redundant Structures - These serial machines (snakes) have excess inputs for a very high level of dexterity and obstacle avoidance capability but require a correspondingly high level of decision making intelligence to operate in real time.

9. Graphical Simulation - In order to design or operate complex robotic structures, their full operational characteristics must be on display with great fidelity to the designer as well as to the machine operator. Training functions (similar to the Link aviation trainer) will become increasingly important for surgeons (micro-surgery), nuclear reactor maintenance, space station operations, etc.

10. Man-Machine Interface - As the technology becomes more complex, a greater need (not less) will develop for a balanced control (or intervention) by man and machine. This will require a much higher level of machine intelligence to obtain the full benefit of the technology for man.

11. Machining Robot - The heart of the factory of the future will require inexpensive generic robots to perform precision light machining operations by direct computer control in order to have a maximum value-added benefit and response to the individual consumer. This requires a complete dynamic and vibration model implemented with feed-forward compensation in real time to make the system electronically rigid.

Level of Funding

The proposed super-robot effort is the most far-reaching program ever envisioned in the field of robotics. Its full level of funding would involve not fewer than 10 faculty and 30 to 50 graduate students with a funding level approaching $3,000,000 per year. Initially, it is proposed to support 3 faculty and 10 graduate students at $900,000 per year. A foundation laboratory is already being developed at The University of Texas at Austin. Specialized access computer hardware for interface with a supercomputer (preferably the ETA 10) would also be necessary with an estimated cost of $200,000.

The Research Team

The present research team has been involved in dynamic model formulation, deflection analysis, real time computation, optimization and design for the past decade and has produced 11 major reports during the past year in this field. The team is establishing a complete robot laboratory to develop next generation technology at the University of Texas, using State of Texas, NSF, and AF resources. About $700,000 will be expended this year on the laboratory.

List of Major Reports for '84-'85 (Co-authored by D. Tesar)

1. "The Design and Analysis of Hybrid Parallel and Serial Robotic Manipulators", Mike Sklar, April 1984.

   Most existing robotic structures are serial in nature (one link, one joint, one link, etc.), in addition to a compact 3-DOF module to act as a wrist. This work extends the general mathematical formulation of the dynamic model by Tesar and Thomas to accommodate mixed serial and parallel structures, especially represented by parallel structural modules in the wrist, elbow, or shoulder locations.

2. "Deflection Prediction for Quasi-Static Serial Manipulators", Allan Fresonke, April 1985.

   The deformations in the serial structure of robots is obtained by representing the six modes of deflection of each link as pseudo joints (6N) which can be added to the N prime mover joints by means of the influence-coefficient method developed by Tesar and Thomas. A very complete model formulation is established to account for end-effector deflection due either to applied external loads or inertia loads. Also, a method is given to compensate for these deflections in order to eliminate their effects and maintain the desired precision of operation even under disturbances.

3. "Kinematic and Dynamic Modeling, Analysis, and Control of Robotic Mechanisms (Via Generalized Coordinate Transformations)", Robert A. Freeman, April 1985.

   As robotic structures become more general, devices such as walking machines (4 and 6 legs with several inputs each), dual arm robots (a total of 12 inputs), and redundant structures such as snakes (more than 6 inputs) must be treated. Their generalized dynamic model formulation (both for serial and parallel structures) is developed in terms of influence coefficients in order allow the designer complete freedom to locate his prime movers in ideal positions in the structure. In addition, formulations are given for a third order description of the dynamic properties of the system as well as a means to mathematically transfer any or all of the prime movers to any location in the structure.

4. "Computer Aided Optimization in the Dynamic Analysis and Parametric Design of Robotic Manipulators", Hsin-Chien Yuan-Chou, August 1985.

Non-linear optimization theory is applied to improved distribution of actuator parameters for robotic manipulators. Formulations on optimal criteria based on vibrations, precision, etc., are established. The minimal number of design parameters (geometry, mass, control, etc.) is 108 to describe a highly non-linear and coupled set of controlling equations. This complexity forces an indirect approach to optimization which is only in the earliest stages of development.

5. "Vibration Analysis and Parametric Identification of Flexible Serial Manipulators", Fariborz Behi, August 1985.

A lumped mass and deformation model described in terms of influence coefficients with pseudo inputs at the principal deflections is used to model the vibratory motion and to predict the frequencies of the lower modes. In addition, modern modal analysis equipment was used to analyze a Cincinnati Milacron T3-776 robot to identify local stiffness and mass parameters and to experimentally verify the vibration model and frequency predictions of the lower modes.

6. "The Design of a Universal Spatial Seven Degree of Freedom Manual Controller for Teleoperator Systems", Mark Bryfogle, August 1985.

Robotic systems used in unstructured task environments such as nuclear reactor maintenance require the full integration of the human operator's judgement and decision making capability. This effort established the design criteria for a kinesthetic force feedback manual controller of one extra degree of freedom as an assemblage of structural modules at the wrist and shoulder. The goal is to enhance the flow of information to and from the operator in real time and to use supervisory techniques to eliminate gross errors, filter jitters, and perfect the global input commands from the operator.

7. "Real-Time Computation of the Influence Coefficient Based Dynamic Modeling Matrices for Improved Manipulator Control", John P. Wander, August 1985.

The complete dynamic model for a 6-DOF serial manipulator is developed in terms of a computational system compatible with array processing. The algorithm operates on a moderate sized array processor (the Analogic AP-500) with a resulting 7.5m sec. cycle time required to represent the most general 6-DOF serial robot structure. Up to 20 DOF systems were tested showing that complex structures can now be operated in real time to make them electronically rigid in terms of existing computational hardware.

8. "Symbolic Programming for Dynamic Modeling of Serial Manipulators", Matthew Reischer, August 1985.

A symbolic Programming Language (AsPL) has been developed to aid in the evaluation and reduction of the controlling dynamic equations for general serial manipulators. This is believed to be the first major effort to create a completely general language structure to treat the dynamic description of robotic systems in terms of generalized matrices (Jacobian, Mass, Coriolis mass, etc.) and to formally organize the development of the controlling equations. Initial results were obtained to determine the best computational resource allocation for various descriptive terms in the dynamic formulation.

9. "Advanced Adaptive Control of General Mechanical Structures and Robotic Manipulators", Sabri Tosunoglu, October 1985.

This adaptive control scheme adjusts the control laws of the prime mover system to reflect the real time condition of the full non-linear and coupled nature of the mass and external load (disturbances) effects on the stability and precision of the system. The claim for global stability is based on a Liapunov analysis. Initial results are obtained on the effects of computational sampling rates and the associated assurance of stability. Preliminary work on criteria for precision of the system's end-effector motion is also described.

10. "Kinematic and Dynamic Analysis of a Six Degree-of-Freedom Parallel Micro-Manipulator", Denise Rill, October 1985.

A 6-DOF parallel structure small motion device (±0.1 inch, ±2 degrees) has been designed as a module weighing 20 pounds and a overall size of about a 7" cylinder 5" high. This module would be placed between the end-plate of the robot and the end-effector to make very small motion corrections much more rapidly than is feasible by the large actuator control system normally found in robots. Influence coefficient analysis has been used to create a dynamic model of this device and to establish design criteria for its most effective operation.

11. "Mathematical Formulations for the Dynamic Phenomena of Generalized Mechanical Structures and Robotic Manipulators", Mark Thomas, October 1985.

This report presents a unified theoretical foundation for the mathematical description of most dynamic phenomena associated with the operation and control of completely general mechanical structures and robots. Emphasis on the generalized serial manipulator leads to the most compact and computationally efficient formulation available for robot system control. Foundation concepts for deformation, parallel structures, higher order control, optimization for design, motion programming, etc., are presented in terms of a powerful notation and unifying theory built on geometric influence coefficients.

II. NEXT GENERATION OF TECHNOLOGY FOR ROBOTICS

BY

DELBERT TESAR

TABLE OF CONTENTS

137

I . APPLICATIONS APPROPRIATE TO AN ADVANCED ROBOTICS TECHNOLOGY

This section of the document is intended to list a representative collection of the most demanding and rewarding unmet applications of robotics. In each case, the application is described in terms of its economic merits, its technical feasibility, and its benefits to the user. Also, in each case, some indication is given as to the necessary technological developments required to satisfy the associated application. One of the more immediate technological gaps is associated with the "open loop" operation of all existing robotic manipulators. Because of this inadequate sensing and real-time compensation based on an inclusive dynamic model, it is impossible to maintain spatial coordinate accuracy (with or without external disturbances and loads). Hence many precision operations at small and large scales (micro-surgery, precision light machining, laser welding, etc.) remain unsatisfied. It also means that off-line programming is normally not possible, such that on-line teaching (while no production occurs) is necessary. This means that the data base cannot directly control the robotic system nor support real time inspection. Hence, special precision assembly, clean room operations, micro-assembly and inspection are less likely candidates for robotics. In batch mode manufacturing, this deficiency means the continued high expense and use of numerous machining jigs—a barrier to the factory of the future.

In other applications, the dexterity and obstacle avoidance of existing systems is inadequate. Many of these systems must work in an obstacle strewn, unstructured environment. Here, special sensing and an advanced machine intelligence must enhance the information sent to the human operator at the man-machine interface to augment his judgment and decision making capacity. Time, frequently, is of the essence so that exceeding human operating speeds is highly desirable. Strategy and planning to deploy friendly forces and strategy identification of unfriendly forces sometimes using incomplete data bases or fuzzy data for military operations has yet to be treated as more than a concept.

INDUSTRIAL AUTOMATION

1. Micro-processing is the spectrum of application of robotics to very small scale industrial operations such as wire soldering of leads to micro-chips, visual inspection and repair of very small assemblies, etc.
2. Complex assemblies involves sequential piece assembly within an obstacle strewn environment where perhaps more than one robot would be necessary (i.e., mounting of a shock absorber on a car).
3. Precision light machining refers to lightly loaded machining tasks in thin stock such as routing, trimming, and deburring while maintaining high tolerance without supporting jigs.
4. Welding in ship structures requires the placement of imprecisely cut thick stock in an egg-crate array and autonomously welding the parts in place.
5. Riveting and deriveting of airframes is the semi-automatic procedure of rivet location, rivet removal, hole inspection and refurbishing, and rivet replacement on airframes with minimum human involvement.

ENERGY SYSTEMS OPERATIONS

6. Nuclear fission reactors could be maintained by robotic systems (especially the steam generator) with minimal occupational radiation exposure and an economic benefit to the nation by 1990 of $1.8 billion/year.

7. Nuclear fusion reactors will require a much higher level and more frequent remote maintenance than fission reactors if their availability is to be 75%.

8. Oil exploration and production on the ocean floor involves maintenance and inspection of the complex ocean floor technologies (valves, pipes, pumps, etc.) in an unstructured environment.

9. Coal production is responsible for 200 deaths/year and considerable cost to the nation due to black lung disease, a dilemma that could be reduced by developing a "manless" coal mine.

10. Nuclear fuel handling and reprocessing has been a long time user of robotics and is now experiencing a new level of technological development at the Oak Ridge National Laboratory.

MILITARY OPERATIONS

11. Remote ocean operations concerns the remote surveillance, personnel retrieval, repair and tactical operations in an unstructured ocean environment.

12. Battlefield operations represents a complex of operations such as surveillance, autonomous tanks, mine removal etc. to remove personnel from the war zone.

13. Maintenance and emergency repair technology is intended to remove technically trained personnel from the war zone as well as make emergency repairs more cost effectively and reliably.

14. Fuel and ammunition handling will reduce logistics problems, increase reliability, and reduce personnel exposure in the war zone associated with the movement and palletizing of materiel.

15. Planning and strategy operations will augment the field commander's decision capacity as the complexity of field operations increase and provide him with an assessment of the strategy of the unfriendly forces.

HUMAN AUGMENTATION AND AGRICULTURE

16. Micro-surgery is intended to augment the precision of the surgeon's motor capacity by a factor of 10 and increase his productive life for operations of the brain, ear, eye, nose and throat including exploratory diagnostics.

17. Prosthetics and orthotics suggests that many partially incapacitated human joints and limbs could be either supported passively or actively to provide improved structural function or they could be replaced by advanced intelligent prosthetics.

18. Agricultural operations associated with non-cereal production are labor intensive and frequently under weather threat. Robotic handling and harvesting equipment could not only reduce costs but also reduce production uncertainties.

19. Accident missions suggest using robotic systems in surveillance, people retrieval, and active threat reduction associated with fires, earthquakes, terrorists, and bomb removal and disabling.

20. Training and service robots are intended to augment humans in education at all levels (truck operation to micro-surgery) with future systems developed for cleaning and maintenance in both public and domestic applications.

INDUSTRIAL AUTOMATION

1. Micro-processing

One of the future opportunities for intelligent machines and robotics is the performance of various precision operations at very small scales. Examples involve the mechanical handling of very small electrical components, wire soldering of leads to micro-chips, visual inspection of very small assemblies, and mechanical or electrical (by laser beams) repairs of imperfect components. Generally, as the complexity of computers, avionics, and precision instrumentation increases (including medical instruments), the need for minaturized systems (micro-robotics) will also increase. Autonomous and teleoperator type systems will be necessary depending on whether the operation can be highly structured or will require the judgement and decision making of a trained human operator. This type of technology will make clean room operations increasingly cost effective and more widespread. In order to achieve this level of technology, a generic class of miniaturized bearings (perhaps jeweled), specialized sensors and encoders, and actuators must be developed. Generally, as the scale is reduced, the relative importance of friction increases such that special antifriction measures will be necessary. Also because tolerances will be so small, programming the system by visual inspection will be much more difficult. Consequently, miniature robotics will require a much higher level of machine intelligence than that found in present generation robotic systems.

2. Complex Assemblies

Most assembly processes currently being performed by robots are especially designed to take into account the present limitations of the robot system. For example, the assembly of an electric motor allows a sequential stacking of components about a vertical axis of symmetry. The insertion of electrial components on a circuit board occurs only on one plane within virtually a perfectly rectangular array. None of these operations must occur in an obstacle strewn environment. Furthermore, most of these steps require relatively little force during the joining stage. Hence, they can be classed as precision, unloaded, unobstructed assemblies.

Future assemblies must treat a much broader range of tasks including force fit assembly, fastener deformation, the joining of heavy components, joining of components by single or repeated impacts, etc. Furthermore, as the assembly progresses, access to the work scene will be obstructed either by jigs or the parts of the unfinished assembly. Consider the difficult task of putting shock absorbers on the suspension of an automobile as one of the type of assemblies that must be addressed in order to expand the market of robotics. This expanded class of assembly task will require precision under load (now universally lacking), a very high level of dexterity, obstacle avoidance routines by an advanced machine intelligence, and a combination of sensors and dynamic modeling in real time to "close the loop". This level of technology will require the most of integration of all mechanical and electrical technologies and will require a concerted research and development effort.

3. Precision Light Machining

Batch mode manufacturing implies numerous light machining tasks where high precision and rapid changeover from one task to the next is necessary. Two example tasks are deburring and trimming of surface panels of aircraft. Because of the large deformations experienced by robotic manipulators operating under these machining loads, jigs or fixtures are used to resist these loads. Of course, the jig must be matched in shape to each part to be finished. In batch manufacturing, literally hundreds of parts (and therefore jigs) are involved - hence, the costs of manufacture and handling of these jigs is very high. Other costs are involved. Programming the robot still relies on step-by-step teaching in the work environment which is very time consuming - a time during which no production is possible. Also, the jig interrupts the flow of information from the unit process to the factory computer data base making the factory of the future impossible.

All of these high costs items for robotic systems could be greatly reduced if precision under load could be achieved. The first need is a complete and accurate parametric model of the industrial robot manipulator (rarely exists and certainly is never used in real time operation of today's robots). Next it is required to make real time (1/30 sec. sampling rate) computer control of the robot in terms of this dynamic model a reality. Also, off-line programming must be developed to make the robot absolutely accurate in world coordinates. Finally, the system must be able to eliminate force induced deflections from the machining process by compensating commands to the actuators. All of these technical objectives will require the best machine intelligence based on the most advanced analytical tools from the mechanical and electrical disciplines. This "closed loop" concept is an essential component of the next generation of robot - i.e., the fly by wire robot which may then be considered "electronically rigid."

4. Welding in Ship Structures

Welding is one of the most important joining processes used in the United States where almost 1,000,000 workers claim to be welders. Automation of welding has gradually taken place by using automatic wire feeding and special seam trackers in conjunction with "tractors" capable of following a straight seam. Hitachi of Japan has implemeted a 20 pass weld of a precision cut joint between two pipe sections with a 1" wall thickness. The seam geometry in many applications is far from straight and for thick weldments (above 3/16") it is difficult to maintain uniform seam spacing or seam alignment. This fact is especially true of welding associated with ship hulls. Also, the ship hull is large and appears to have an "egg crate" geometry in much of its multiple wall and multiple cell configuration. This reality makes mobility, dexterity, obstacle avoidance, superior sensing, and high precision essential to a ship welding system. Beyond this, excessive teaching time for such complex geometries becomes a dominant problem of the existing robotic technology. This is compounded by the fact that shipbuilding is characterized by its variety of small batch operations (often unique assemblies) where programming time can represent as much as 90% of the total processing time.

In order to eliminate most of these problems, the robot welding system must be driven from a data base of the ship component being welded. Reference points on the tacked assembly can be used to automatically place the workpiece in the coordinate system of the robot. Then if the robot is absolutely accurate and if the welding process is monitored with adequate sensors (such as vision), the welding procedure can be achieved with virtually no teaching time. Of course the data base must tell

the robot where the "obstacles" of the incomplete assembly are and how they change as progress is made. The robot intelligence must be capable of avoiding these obstacles without human intervention. It is also feasible that one robot could be used to place parts in the assembly while the second robot performs the necessary weld.

5. Riveting and Deriveting of Airframes

Hundreds of thousands of rivets are used to assemble the airframe of modern airplanes. Even though special hand held riveters are used effectively, they require a great deal of heavy labor. It has been frequently suggested that robots be used to hold the riveting unit during the riveting process. In order to perform this task, the riveter must be perpendicular to the surface and perfectly aligned with the rivet hole. The surface geometry and rivet array forms a complex spatial geometry which demands that the robot have generic motion capability. Because the riveter is heavy the gravity forces will cause significant deflections. These deflections and those due to riveting forces must be compensated for by an active machine intelligence capable of positioning the riveter accurately in coordinates attached to the airframe. Then and only then can the data base control the robot directly. Otherwise, the robot must be calibrated and taught for each section of every airplane one at a time. Such teaching effort would consume more time and higher expense than the previous manual operations.

Airplanes placed on aircraft carriers experience sea salt corrosion of the rivets making it necessary to derivet the airframe. All of the above requirements apply. In this case, it is also necessary to accurately drill out the old rivet. Teaching the robot is clearly impractical because of all the uncertainties and load variations. An advanced form of closed loop control of a precision robot combined with computer vision could make self-calibration of the robot relative to the airframe feasible so long as the data base for that air frame were available. Such a system would make it possible to derivet the plane on board the carrier or in a remote field operations shop making repair logistics much more economical.

ENERGY SYSTEMS OPERATIONS

6. Nuclear Fission Reactors

There are major economic losses associated with the critical path down time and occupational radiation exposure (ORE) associated with Maintenance, Testing, and Inspection (MTI) of nuclear fission reactors. The total cost of these operations in 1980 was an average of $14 million per plant. This represents an approximate $1 billion cost to the nation. When using steam generator maintenance as a vehicle for analysis, a 70% savings is predicted; i.e., a total national savings of $700 million per year. By 1990, the projected savings would reach $1.8 billion per year. If ORE limits were lowered by a factor of 5 (as has been suggested), these numbers would all grow by approximately 75%.

The present level of manipulator technology is insufficient to perform most of the needed maintenance tasks in a successful and time efficient manner. The present and near term reactor was not designed for remote maintenance thus making the need for a generic maintenance system more pressing. This generic system must be capable of performing a series of up to 25 distinct operations in an obstacle strewn environment with the characteristics of a portable machine shop. The PWR steam generator and the BWR valve have been isolated for immediate application of robotics. Specific component technologies which must be addressed to meet this need

are multiple task capability, a high level of machine intelligence, man-machine interface, dexterity and obstacle avoidance, precision and load capacity, and portability and mobility.

7. Nuclear Fusion Reactors

Plant availability of fusion reactors will be closely linked to the effectiveness of the remote maintenance technology to be employed. Many experts believe that rather great technical advances in the technology will be necessary. Princeton's TFTR for example, is planning for an availability of about 2%, but even this has not yet been demonstrated with current technology. The Fusion Engineering Device (FED) of Oak Ridge has set a 50% availability goal. The objective of the planned Starfire fusion utility plant must be 75% availability.

The fusion reactor represents a highly uncertain environment, thus calling for a completely general remote system with advanced machine intelligence and man-machine interface. The most difficult remote task will be the handling of the 400 ton shield sector. The shield frame must be precisely positioned, cut, and welded. The required manipulators must handle loads up to 500 lbs. The large loads combined with the large arm dimensions, place extreme demands on the technology where precision requirements are very high. These projected requirements will not be met by evolutionary development expected from industrial laboratories. A much more complete understanding of the complex geometry and dynamics of the manipulator system is needed to produce these larger maintenance devices.

8. Oil Exploration and Production on the Ocean Floor

At the present time, the majority of operations is carried out by human divers, representing an extremely dangerous activity. The present trend towards deeper open-water wells represents an even more dangerous operation. Operations below 450 meters must be carried out by Remotely Operated Vehicles (ROVs). There are two major limitations in the use of ROVs. Present vehicles are controlled by tethered lines, which become tangled and severely limit mobility. Also, due to the lack of advanced manipulator technology, only simple tasks such as monitoring and inspection may be carried out using ROVs. If the offshore oil industry plans to reach greater depths, an emphasis on improved remote technology is necessary. During the 5 year period (1975-1979), approximately 100 wells were drilled at depths over 1000 ft. Subsea production systems (underwater wells) are used at depths of 2000 ft. The investment for a total operational platform may easily exceed $1 billion.

ROVs can operate at 1/10 the cost of human diver systems primarily because of reduced support facilities and personnel. Also, descents and ascents require long periods and in bad weather significantly increase risks. There are over 140 tethered ROVs in use today. Approximately 30 vehicles have manipulator capabilities. Most of these manipulators cannot perform the highly useful but difficult tasks of welding, cutting, bolting, etc. As the production system evolves it will become more modular (i.e., valve modules, pipe joint modules, etc.) enabling more rapid and reliable replacement by remote systems technology. A comparative increase in the effectiveness of this remote systems technology would yield major economic benefits.

9. Coal Production

It appears that significant economic benefit and miner safety would result with the use of remote systems technology in deep underground coal mines. Almost no such remote technology is presently used in underground mines. In other words, very little progress has been made to make coal production possible without the direct involvement of human operators - i.e., the manless mine is a distant possibility. The potential areas for automation and robotics are roof support, material handling, fire control and rescue, and surveillance. Roof-bolting is the most dangerous activity in underground mining, causing 50% of all mine fatalities. The mining industry is five times more dangerous than the average U.S. industry, with about 200 deaths occurring each year. The cost of injuries based on 1974 data, to all sectors of society (industry, personnel, health agencies) was $34 million.

Automation of the coal mine will involve the implementation of several dedicated machines. For example, the continuous wall system represents an annual return on investment from 15 to 23% depending upon the level of automation implemented. The use of remote systems technology will be vastly improved when these dedicated machines will be modularized making maintenance by module replacement feasible. Then robotic maintenance systems can be developed with high mobility for this task as well as that of monitoring and surveillance. This combination would make the manless mine possible and significantly reduce the costs associated with maintaining an environment which is suitable and safe for the coal miner.

10. Nuclear Fuel Handling and Reprocessing

Nuclear fuel handling and processing was the first application of robotic manipulators. The technology was developed from 1945 to 1965 at the Argonne National Laboratory. Many hot cell manipulators are used for this purpose today. An evolving application which is now being pursued is fuel reprocessing at the Oak Ridge National Laboratory. There, a new generation fuel reprocessing plant is being designed for implementation late in the 1980's. Such a system which is made up of literally thousands of components that could fail must be maintained remotely once it starts operation because of the high internal radiation levels. One of the first requirements of the maintenance system is extraordinarily high reliability. Not only must the plant be modular and structured for maintenance, so must be the robotic systems used for maintenance. The robotic system must be moved on tracks anywhere in the plant making effective communication with an external data base difficult. Because literally hundreds of maintenance tasks are involved, a human operator must supervise or manually control the task performance through a highly developed man-machine interface. A new generation of robotic manipulator is being developed for this application as are special interface technologies to the operator (visual graphics, voice commands, force feedback, etc.). The goal is to use an advanced machine intelligence to reduce the burden on the human operator by automating as many operations as posseble.

MILITARY OPERATIONS

11. Remote Ocean Operations

The Navy has established a program to meet both tactical and strategic objectives. All indications are that much of this work remains at the conceptual or exploratory stage. One of the first objectives is to perform search and identification of sub-sea systems. Beyond this, the goal is to perform retrieval functions

of lost hardware or stricken submarines. Finally, the most demanding task will be associated with anti-mine and anti-submarine activity. In the strategic sense, the laying and maintenance of under-water communication cables and power transmission lines is a very high priority. It is clear, that the operating environment, especially for maintenance, is unstructured and could be necessary at any depth of the ocean. For example, the well understood task of underwater hyperbaric welding can not be performed by the present technology. All present systems are tethered ROV's or self-contained diving chambers (i.e., specialized submarines). Only one of these systems (the ROV ORCA) presently offers the necessary feature of force feedback to the operator or the controlling computer.

Generally, this range of applications will require the most advanced generic robotic manipulator technology possible. A number of the systems must be untethered to be effective. The dexterity, sensing, and precision of these systems must be very well integrated. The man-machine interface question is also of the highest possible importance due to the unstructured nature of the task spectrum. Finally, these systems must operate with extraordinary reliability with time as the essence. Overall, robotic technology within the ocean will require the best of all component and system technologies.

12. Battlefield Operations

The primary objective is to perform rapid advance maneuvers with minimum exposure to unfriendly forces. Several dedicated units such as autonomous mine detection and disabling vehicles, autonomous offensive tanks, etc., will be essential. In addition, battlefield communication line networks must be established and maintained. These functions can be performed autonomously only if a collection of sensors are developed (acoustic, optical, electromagnetic, tactile, etc.) which are field hardened and highly reliable. In addition, computational vision based on stored object knowledge must be capable of recognizing objects. All of this sensory information must be forwarded to an on-board central processor whose machine intelligence is capable of reasoning and developing a strategy for action. This strategy must be carried out with high reliability to benefit friendly forces.

Autonomous systems tend to be operations specific. Therefore, alternatives which allow human intervention by teleoperation should be carefully considered. In this case, precision heavy duty robotic manipulators may become an essential device for disabling unique multi-purpose and mobile mine fields (under enemy control). The same teleoperator system could prove highly beneficial in laying mine fields in or behind enemy lines. Or, it may prove feasible to develop an autonomous roving mine field which would be targeted against unfriendly forces. Overall, the question of artificial intelligence appears to represent a technological gap which must be met for this application.

13. Maintenance and Emergency Repair

One of the basic realities of modern military materiel is that it is complex and must be continuously monitored and maintained. This is especially true of the increased use of electronic components such as avionics. Maintenance and emergency repair requires a highly trained practitioner in order to diagnose and correct malfunctions. Unfortunately, the most pressing (and valuable) maintenance operations occur in battle zones or in remote locations such as on board aircraft carriers. This means that many highly trained personnel are exposed to unfriendly forces.

Two steps can be taken to reduce the exposure to technically trained personnel while at the same time making emergency repairs much more reliably. One of these steps has already been established within the field of avionics and that is self-contained modules that are easily interchanged. This design philosophy must be used on the mechanical system as well. The second step is to make maintenance through teleoperation feasible by deploying a generic maintenance robot system having a precision dexterous manipulator with force feedback to a remote station where the operator works with a multi-faceted man-machine interface. Because emergency repairs may be required due to damage from enemy fire, the task spectrum must be considered as unstructured thus requiring a high level of human judgement and decision making to make the maintenance repair as reliable as possible. Here, the technological gap appears to be a generic precision, mobile robotic manipulator with some machine intelligence supported by a superior man-machine interface.

14. Fuel and Ammunition Handling

In the deployment of tactical units, the fuel and ammunition zones are the most likely targets of unfriendly forces and when hit cause potentially severe destruction. Hence, minimizing personnel in this zone would be a high priority. In addition, during engagement rapid loading of fuel and ammunition is a very clear necessity. The major time element is associated with ammunition loading of such units as tanks. It presently takes 3 to 4 hours to load a tank with its full complement of rounds. It is desired to reduce this to one hour - therefore, potentially doubling the availability of the tank. A recent example of automatically loading an A-10 anti-tank aircraft in 8 minutes relative to a period of 3 hours for manual operation shows that a truly integrated system can reduce loading time by as much as a factor of 20 times.

The envisioned system would employ a heavy duty robotic manipulator to semi-automatically depallet the ammunition and pass it to a transfer device at the manway of the tank. The transfer device would lower the round to a reference rack in the tank from which a dexterous robotic manipulator would take the round. The internal manipulator would remove any "duds" as a first step in the return cycle and take all incoming rounds and automatically palletize them in the tank. The internal manipulator would also be able to take the rounds out of the tank pallets and insert them into the tank gun barrel. This internal system would then make one crew member redundant. The external manipulator would be rather large and somewhat mobile on its own platform. The biggest technological gap would likely be associated with the highly dexterous, high load capacity, precision manipulator (internal to the tank) which should be operated autonomously, especially during maneuvers. Such inclusive technology will require very high integration of some immature but emerging component technologies.

15. Planning and Strategy Operations

Today, planning and strategy development is becoming increasingly important to assist personnel in making short term and long term decisions about troop and materiel movement and deployment. As the number of distinct and sophisticated fighting elements (roving mines, autonomous tanks, controlled electronic barriers, etc.) deployed by unfriendly forces expands, the need for more complicated and more rapid decision making becomes critical. In addition to these managed "obstacles", there are terrain obstacles such as boulders, trees, swamps, and rivers. The obstacle strewn environment is one of the unsolved planning problems facing the robotics research community. Presently, the problem is partially solved by trained personnel on-board the dedicated vehicles (tanks, supply trucks, etc.).

In the near term these tasks could be taken over by teleoperation if the control task is relatively simple and no on-board activity demands human activity. As the task becomes more implicit, because of invisible managed obstacles by unfriendly forces, it becomes more difficult to adequately train the large numbers of personnel required in the field. Hence, in the long term, on-board computers will be required to provide planning and navigation. Planning involves data acquisiton (perhaps fuzzy) to augment an existing data base, reasoning among alternatives (serially or in parallel), accounting for coupling among on-going actions (spatial reasoning) and in process control through monitoring and time efficient up-dating of the planned operation. Navigation accounts for the existing system configuration (geometric status) among stationary or managed obstacles (avoidance strategies) to develop routes (optimal paths) by means of a global awareness. Several important technological gaps exist. Primarily difficulties arise from fuzzy acquired data or incomplete data bases. Route planning involves one of the most demanding of all optimization problems if the obstacles are numerous or managed effectively by unfriendly forces. Finally, strategy identification through analysis of the "movement of managed obstacles" would prove invaluable to decision makers in the field.

HUMAN AUGMENTATION AND AGRICULTURE

16. Micro-Surgery

Micro-surgery involves the use of a microscope to enhance by a factor of ten the vision of the surgeon. At this point in time, this has been a major advantage in the fields of eye, ear, throat, and brain operations. In addition, much research now involves work with single cells and requires the best available precision in mechanical operations. The primary need demonstrated by this activity is to augment the human operator's motor capacity (i.e., the surgeon's) in order to complement his enhanced visual capacity perhaps by an order of magnitude (by a factor of 10). One of the goals of this type of system is to lengthen the productive life of the surgeon. The other immediate goals of improved precision can be achieved by filtering jitters and oscillations from muscles that over-react under tension and by changing scales of the operation through computer enhancement. Thus far, little has been done to satisfy this need with a flexible all-purpose (generic) system. Miniaturized robotic systems have not experienced much development to-date. Three component technologies are important to this application. First, as the system becomes smaller, friction becomes relatively more important making special frictionless bearings an imperative. Secondly, because the tasks are at such a small scale, special miniaturized force sensors must be implemented to keep the surgeon in close awareness of the operation. Thirdly, the surgeon must work through a superior transparent man-machine interface in order to make him not only comfortable but fully in charge of the process. These component technologies are very demanding, and thus far poorly developed. It is estimated that a major team of researchers would require a decade to implement this technology in clinical operations.

17. Prosthetics and Orthotics

Despite the fact that no method for preventing or curing the many arthritides afflicting man have been found, significant advancements have been made in the past decade toward providing an adequate substitute for destroyed joints. The major improvement in the care of arthritics has been the development of internal prostheses. A nominal number of passive orthotic devices have been developed to act essentially as braces for the weakened human structural system. A major opportunity

now exists because of development in robotics and micro-electronics to develop active aids to the human system where the muscular activity is diminished or atrophied. Or it may be possible to replace the function totally by an alternative device. In the first case, the kinematics of each structural element (the knee, ankle, shoulder, etc.) must be examined in vivo to exactly duplicate its residual motion. This data can be used to design and demonstrate a class of actively driven orthotic structures to replace or supplement the existing weakened muscular activity. Such devices could be extremely valuable in training or strengthening muscles that have experienced trauma. In the second case, total replacement by means of an actively powered device (prosthesis) may be necessary. Here remaining muscles can be trained to generate electrical signals to be interpreted by a microprocessor which then would control the actuators of the device. A frequent objective to assist quadraplegics is to implement articulated wheel chairs or roving robot servants. All of these systems must be designed for the lowest possible price, be as light weight as feasible, and be exceptionally reliable. Also, in every case the man-machine interface must be carefully matched to the individual be it kinesthetic, voice, or visual. Indications are that this technology could be pursued vigorously today and satisfying results would be expected.

18. Agricultural Operations

One of the primary problems facing many agricultural operations is the relatively high cost of labor intensive tasks associated with such functions as fruit harvesting. The economic loss of inefficient or untimely harvesting (the weather threat) can be devastating. The alternative pursued today is total plant harvesting where specially bred plants produce fruit that ripens simultaneously thus allowing the plant to be destroyed during harvesting (i.e., as with tomatoes). This may lead to both economic and quality compromises. Hence, it is proposed to demonstrate a new class of agricultural system which is capable of independent action depending on the requirements of the unit operation. This may be illustrated by the example of citrus harvesting. In this case, the ripe fruit can be identified by its unique color (orange) relative to a dark green background. This identification can now be accomplished by computerized vision which would identify the fruit and provide data to the central processor to quantify the location of the fruit. On this basis, a robot arm could be instructed to pick the fruit (a fact confirmed by touch sensors). It is recommended that an array (perhaps 20) of inexpensive modular robot arms be used to perform this function, each moving with relatively low speed. If one failed, it could be temporarily removed without shutting down the rest of the harvesting system. Similar developments could be pursued in greenhouse, packing house, warehouse, cut flower, and packaging operations. It now appears feasible to create a whole new class of technology specifically suited to non-cereal grain agriculture.

19. Accident Missions

One of the unmet opportunities for robotics systems involves rescue and surveillance activity associated with accidents such as earthquakes, fires, terrorist bombs, etc. Recently, New York City and several other cities have employed a roving robotic system to remove or disable terrorist bombs. Each year, many police personnel are injured or killed from bomb explosions. Also, in pursuit of dangerous criminals, police frequently have to expose themselves to attack during surveillance or apprehension. In the case of fires, firemen must make every effort to determine the whereabouts of trapped individuals in an on-going fire. Fire resistant robots could be of real value in this surveillance function as well as

providing sustenance (food, water, oxygen) or protective cover (fire retardent clothes) to those trapped. Earthquakes require special material handling needs to uncover persons trapped below ruins.

All of these applications require various levels of mobility, sensing, and on-board intelligence. The robotic manipulator itself may be either simple or heavy duty depending on the application. In disabling bombs, the dexterity and visual and force feedback to the operator at a remote location will have to be of the highest order so that accidental activation of the bomb triggering device can be prevented. For fires, it may be necessary to have the device climb the sides of buildings in order to gain access to upper stories of buildings. The robot could carry a lightweight cable which could be anchored on both ends. Then a powered trolley could travel along the cable to rescue trapped individuals. It appears that accident mission robots could have an enormous impact in life threatening accident situations.

20. Training and Service Robots

A future opportunity for robotic technology will involve human augmentation in the broadest possible sense. One of these functions will involve training objectives as now being demonstrated in ground based training systems for the beginning pilot. Here, the system duplicates the flying cockpit environment as accurately as possible including visual and motion feedback to the operator. Similar training systems will be of high economic importance where the actual system (say the operation of trains, large trucks, ships, nuclear reactors, surgery, etc.) is either too expensive or too dangerous to duplicate in reality. In educational institutions, at all levels, robotic technology will be used to enhance functional learning (as now being gained from computer games). Presently, only the simplest digital or analog interface is being used. As an inexpensive man-machine interface becomes more universal, this educational opportunity will rapidly expand.

Service robots have long been envisioned by science fiction and a few elementary examples have begun to appear. None of these systems are known to perform useful work economically. The cost of such systems could easily exceed $100,000. If one considers the functions that would be attributed to a "housebot" one quickly realizes how many unique operations would be necessary. It is conceptually possible to create an autonomous robot vacuum cleaner. This device would carry rechargeable batteries (to be automatically plugged in on demand), be highly mobile, and be able to plan a complete traverse of open floor space while avoiding all obstacles. It appears such a device would have several simply articulated vacuum arms enabling access below furniture and in narrow volumes between obstacles. Eventually, this market will be met but, in the near term, only specialized systems of high value (supermarket floor cleaners) should be attempted.

II. MATRIX OF COMPONENT TECHNOLOGIES FOR ROBOTIC SYSTEMS

The integration of numerous technologies is one of the fundamental realities of robotics (or more generally intelligent machines). Often significant progress in the system development will occur after a breakthrough in a component technology. Hence, except for exceptionally large research facilities, most research efforts will pursue a few component technologies and look to the manufacturer to do the system integration and development. The following 14 component technologies are given to cover the broad spectrum represented by robotics. Each of these component technologies will be described briefly in the next few pages.

1. The structural geometry of the robot, its design and operation for determination of its workspace, reach, dexterity, obstacle avoidance, etc.
2. Structural dynamics of robot systems for modeling of robot dynamic and vibration phenomena for purposes of design and improved operation.
3. Prime movers are the muscles of the manipulator whose precision of operation is dependent on their response and resolution.
4. Actuator modules involves the structural integration of prime movers into modules of 1, 2, or 3 degrees of freedom which can be assembled into a robotic system.
5. End-effectors are the interface hardware and software to perform the handling, inspection, machining, etc. task of the robot; they may include special touch and force sensors.
6. Graphics/CAD of robot phenomena to enhance interactive design and optimization in complex manufacturing environments.
7. Sensor technology is essential to the existence of an intelligent machine so that it is aware of its own existence and process parameters associated with its operation (manufacturing, maintenance, etc.).
8. Vision is the specialized sensor capable by computer enhancement of rapidly digitizing the physical environment of the robot allowing for comprehensive planning and strategic operation.
9. Artificial intelligence structures the decision making process for multi-layered phenomena in the robot system.
10. Intelligent control involves the layered implementation of various control strategies on global and local objectives.
11. Software modules implies the compact and hardened packaging of frequently used algorithms and their specialized chip assemblies.
12. Computer architecture involves the assemblage of serial and parallel processors capable of treating multi-faceted computational tasks within the concept of real-time operation of the system.
13. Communication interfaces involves the structural distribution of operational decisions and data reduction and transfer of the sensor signals among the various components and layers of the total system.
14. Man-machine interface allows direct human communication with the intelligent robot to allow human augmentation in unstructured task applications (micro-surgery, nuclear reactor maintenance, etc.).

TABLE 3: ESTIMATES OF LONG TERM IMPORTANCE OF COMPONENT TECHNOLOGIES FOR VARIOUS APPLICATIONS

Estimates of Importance of Robotic Technology for Various Applications / Robotic Component Technology	Average	Micro-Processing	Assembly	Light Machining	Welding	Riveting	Fission Reactors	Fusion Reactors	Oil Production	Coal Production	Fuel Handling	Ocean Operations	Battlefield Op.	Emergency Repair	Ammunition Handlg.	Strategy Develop.	Micro-Surgery	Prosthetics	Agriculture	Accident Missions	Service Robots
1. Geometry (13)	3.8	3	4	3	5	4	8	8	4	3	3	6	5	4	4	-	4	2	2	3	2
2. Dynamics (14)	3.4	3	5	5	4	4	4	4	4	3	3	4	4	4	5	-	3	2	4	2	2
3. Prime Movers (11)	4.3	4	3	8	4	6	6	6	5	4	5	6	6	5	5	-	3	2	3	3	2
4. Actuator Modules (10)	4.3	3	4	4	4	7	7	7	5	4	5	7	6	5	5	-	3	2	3	3	3
5. End Effectors (9)	4.4	3	4	2	2	3	5	5	7	4	4	7	6	6	4	-	4	2	4	7	7
6. Graphics/ CAD (12)	3.9	4	4	3	5	5	6	6	4	3	4	5	6	4	4	4	4	2	2	2	2
7. Sensor Technology (5)	4.7	7	4	6	5	5	5	6	5	4	4	6	5	5	3	-	6	4	8	4	4
8. Vision (2)	5.8	9	7	4	5	8	5	5	6	4	3	8	7	4	5	-	6	-	5	3	8
9. Artificial (4) Intelligence	4.8	5	4	4	4	4	5	5	5	4	3	6	6	4	4	10	3	4	4	6	6
10. Intelligent Control (7)	4.7	7	7	8	6	7	6	6	5	4	4	4	5	4	5	-	5	3	3	4	4
11. Software Modules (8)	4.4	5	4	6	4	5	4	4	5	4	3	4	4	4	4	8	5	4	3	3	3
12. Computer (3) Architect.	5.1	5	6	10	7	8	4	4	5	4	4	5	8	3	4	8	4	2	4	3	5
13. Communication Interfaces (6)	4.7	4	4	5	5	6	5	5	5	4	4	5	8	6	5	-	4	2	4	4	4
14. Man-Machine Interface (1)	6.4	5	3	3	5	6	10	10	8	5	8	8	4	8	3	-	10	10	8	8	7

TABLE 4: VALUE FOR NEAR TERM IMPLEMENTATION OF COMPONENT TECHNOLOGIES

Robotic Component Technology	Normalized Availability	Average for all Applications	Industrial Automation	Energy Systems	Military Operations	Human Augmentation and Agriculture
1. Geometry	1.0	3.8	3.8	5.2	4.8	2.6
2. Dynamics	0.3	1.0	1.3	1.1	1.3	0.8
3. Prime Movers	0.7	3.0	3.5	3.6	3.8	1.8
4. Actuator Modules	0.4	1.7	1.8	2.2	2.3	1.1
5. End-Effectors	0.3	1.3	0.8	1.5	1.7	1.4
6. Graphics/CAD	0.7	2.7	2.9	2.9	3.2	1.7
7. Sensor Technology	0.6	2.8	3.2	2.9	2.9	3.1
8. Vision	0.2	1.2	1.3	0.9	1.2	1.1
9. Artificial Intelligence	0.2	1.0	0.8	0.9	1.2	0.9
10. Intelligent Control	0.3	1.4	2.1	1.5	1.3	1.1
11. Software Modules	0.1	0.4	0.5	0.4	0.5	0.4
12. Computer Architect.	0.4	2.0	2.9	1.7	2.2	1.4
13. Communicat. Interfaces	0.5	2.3	2.4	2.3	3.0	1.8
14. Man-Machine Interface	0.5	3.2	2.2	4.1	2.9	4.2

200

Of course, all of these component technologies are of primary importance to the implementation of robotics to this spectrum of applications. Nonetheless, a great deal can be learned by ranking the technologies with respect to their near term and long term relative significance. The long term importance of a component technology should act as a guide to the relative emphasis in basic research among the various technologies. By comparison, the near term value of a component technology should provide an indication of the relative development effort now likely to result in the best short term "pay-off" in actual application. The results of an attempt to quantify these two levels of significance are given in the following partial tabulation.

Normalized Long Term Component Importance		Normalized Near Term Component Value	
Man-Machine Interface	10.0	Structural Geometry	10.0
Vision	9.0	Man-Machine Interface	8.5
Computer Architecture	8.0	Prime Movers	8.0
Artificial Intelligence	7.5	Sensor Technology	7.4
Sensor Technology	7.3	Graphics/ CAD	7.1
Intelligent Control	7.3	Communication Interfaces	6.0
Communication Interfaces	7.3	Computer Architecture	5.3

Note that for the two application groups, military operations and energy systems, the two component technologies, actuator modules and end effectors, show high long term significance.

The difference between the near term and long term rankings is due to the fact that the technologies are not uniformly available in the near term where it is assumed that they will have the same availability in the long term. In this case structural geometry is thought to be 50% available, prime movers and graphics/CAD at 35%, while vision and artificial intelligence are considered to have reached only 10% of their real potential.

1. Structural Geometry

The analytical tools to treat the operation and design of the geometric dimensions of robot arms has been found to be one of the most complex problems associated with robotics. The cartesian robot contains no fixed dimensional parameters. Many present dexterous arms (similar in proportion to the human arm) contain two fixed dimensions. The most general 6 degree of freedom arm would contain 18 design parameters all of which should be evaluated to enhance the reach, dexterity, obstacle avoidance, etc. potential of robot arms. Recently, researchers have shown that the complex mathematical control equations may fail frequently and cause disconcerting disruption in the smooth or precise operation of the arm. Future arms will be a balance among the number of degrees of freedom (redundancy of 2 to make an 8 DOF arm) and the level of complexity in the geometry and the associated planning and control algorithms. Almost all existing arms are now serial devices (one link, one joint, one link, etc.). Future geometry will involve the study of parallel structures for enhanced precision and load capacity. The scale of these devices could become very small (miniature manipulators) putting increased demands on the analytical theory and design methodology. Finally, two or more robots could work together to perform an assembly task (say welding). In this case, what is their common workspace, dexterity, and operating region without mathematical uncertainties or special locking configurations? What is the desirable balance of complexity among the interacting arms?

2. Structural Dynamics

Most existing industrial manipulator arms are very flexible and easily deform under load (from 0.2" to 0.4") and respond to simple hand shaking at frequencies less than 10 CPS which means that their fastest cyclic speed would not be better than 30 RPM (compare with most packaging machinery at 300 RPM and some textile machines at 3000 RPM). The associated deformation may be the result of dynamics of the system (usually known) or they may originate from the task operation (routing, force fit assembly, deburring, etc.) which are usually unknown. Many of these future applications of robotic manipulators will require a high level of precision under large load variations. Today, all manipulator systems operate open-loop where neither the dynamics nor the external loads are accounted for. The barrier to meeting this fundamental objective is the ability to create the model in real-time (say about 30 hz). Having the model in real-time would enable the compensation for the system deformations and predicted improvement of precision under load by a factor of 10. As this technology becomes available, more robust control strategy will be implemented to allow lighter weight structures (especially desirable in serial arms). Also, as improved dynamic control occurs, redundant degrees of freedom will be used to enhance controllability. Alternatively, the dynamic modeling could be made more accessible to real-time operation if parallel structures were used. Associated with this activity is the dynamic programming of the end-effector motion to reduce command shock induced oscillations. This objective is closely related to the desire for high speed "slewing and touching" in minimum time. None of this activity can move forward without accurate parameter identification for the link masses, link deformations, actuator control circuit parameters, etc. As many as 130 parameters are involved. Hence, it will be essential to develop design tools and criteria for these lighter and faster arms.

3. Prime Movers

The muscles or energy sources which move the manipulator arm are the prime movers of the system. These components may be electric, hydraulic, or pneumatic. Electric prime movers are increasingly more common. Because of their inherently low load capacity, they almost always require mechanical force amplifiers in the form of gear trains or metal tapes. These amplifiers all add weight, compliance, and backlash, and they increase maintenance and reduce reliability. Hydraulic prime movers, although powerful, exhibit limitations such as fluid leakage (critical in some clean room operations), sensitivity to dirt in the fluid passing through delicate servo-valves, stiction, and variable bulk modulus in the fluid circuit. Pneumatic actuators are inordinately "soft" and very difficult to control for precision positioning under load.

New electric prime movers are appearing (based on rare earth materials) with increasing load capacity and therefore reducing the critical parameter of weight. Amorphous materials (powder metallurgy) may significantly reduce hysteresis losses having the same effect. Better control through PWM of DC motors based on V-MOS technology and hybrid implementation of digital and analog designs should provide enhanced load capacity, dynamic response, and resolution. Antagonistic impulse control circuits may soon be developed with "cross-firing" to further improve positional resolution. Miniaturized prime movers are one of the critical unmet needs required to drive improved robotic hands or micro-manipulators suitable for micro-surgery, micro-assembly, and small scale inspection and maintenance. At this scale, positional resolution degrades rapidly due to the increased relative significance of friction.

4. Actuator Modules

Modularity of the prime mover and its surrounding physical structure is perceived as a major opportunity to reduce the 6 to 7 year design-to-market cycle time now required for new generations of robotic manipulators. These modules (or building blocks) would be a series of 1, 2, or 3 degree-of-freedom (DOF) units which could be assembled rapidly by a designer to respond to the requirements of a given application. Such modularity would do a great deal to increase the breadth and rapid diffusion of robotic systems.

Most actuators presently being used in manipulators are off-the-shelf prime movers not specifically designed for precision control of large coupled motions as occur in robots. This approach does not lead to an optimum balance between the best characteristics of the prime mover and the physical structure of the system. Presently, many actuators are too heavy, have poor response times to commands, generate backlash inaccuracies, have poor resolution, are not stiff under load, and do not contain any local intelligence. The next generation of robot must be constructed from a large class of near optimum actuator modules which contain their own sub-systems for sensing and intelligence. These modules must be rapidly scaled (small and large sizes) with standard physical and software interfaces for effortless assembly. Enhanced maintenance due to this modular design is an obvious benefit. This approach is the primary reason that the application of the modular micro-chip is so widespread.

A manufacturer has recently announced a 3 DOF hydraulic wrist. Cincinnati Milacron has aggressively implemented their three-roll wrist. A Japanese painting robot uses a sophisticated linkage based 3 DOF wrist of high dexterity.

The human system is composed of a 3 DOF shoulder, a 2 DOF ankle, a 3 DOF wrist and forearm, a 2 DOF knuckle, and a 3 DOF hip. These systems are capable of high positional resolution because of muscular antagonism, therefore eliminating backlash. Friction at very small scales can be reduced by using anti-friction ceramic bearings. Parallel linkage structure can be used in the module to create very high stiffness with low weight. Hence, it can be argued that the next generation of robotic system will come a great deal sooner if a major thrust for structural modules were pursued.

5. End-Effectors

End-effectors are the tools attached to the end of the manipulator arm to perform specialized functions such as welding, drilling, locking or unlocking bolt assemblies, etc. Frequently, specialized tools must be interchanged, a process which must be time efficient and very reliable. Some end-effectors are multi-purpose devices in the same sense that the human hand is able to hold a hammer, screwdriver, or other handtool. Generally, the complexity of the terminal device is an inverse function of the complexity (or dexterity) of the arm. As the technology matures, it is expected that general purpose terminal devices (hands) will reduce demands for versatility on the manipulator arm. I.e., small end-effector motions (in the form of a 3 DOF to 6 DOF micro-manipulator) will make large system motion less necessary.

The normal medium size gripper of today is a simple pair of parallel fingers capable of holding a 5" weight of 10 lbs. Generally, these devices are clumsy and require excess maneuverability to grasp a generic object. Frequently, they incorporate some elementary force and proximity sensing. Specialized end-effectors for drilling, sanding, painting, etc. will continue to be developed. All indications are that a new generic hand is required to reduce the number of special tools necessary to perform a range of unstructured tasks. This hand should have 3 or more compliant coordinated fingers of medium dexterity with good incremental force sensitivity capable of grasping and orienting an arbitrary object in space. The power source and intelligence for this generic hand should be contained within the unit itself because of the difficulty of passing control forces through the wrist of the manipulator. Leakage of hydraulic fluids would limit the usefulness of such a hand. Hence, miniaturized prime movers must be developed for this application. The fingers for this generic hand should employ a robust, low hysteresis touch sensor with 1 gram sensitivity and a dynamic range of 1000 to 1. The desired resolution would approach 1000 points/in.^{2} Preferably, the sensor would process this force data locally at the sampling rate of 100 hz. Once the technology for such a hand has been demonstrated, it will be necessary to fill out the spectrum between it and the specialized devices prevalent today.

6. Graphics/CAD

Because of the generality of motion during operation and the large number of system design parameters, the design of manipulators is an expensive, time consuming and challenging task. The magnitude of this task can be illustrated by noting that a generic six degree of freedom serial manipulator can have as many as 18 geometric parameters, 60 mass parameters, and 42 stiffness parameters along with 12 or more actuator parameters. The design and development of such a generic structure can cost millions of dollars (the space shuttle manipulator cost $100,000,000 to develop). As requirements for precision

operation, cyclic speed, and external loads increase, the ability to meet complex design objectives becomes more critical. In order to provide enhanced system design and expand the designer's understanding of and control over the design process, it is essential to utilize rapidly growing computer capabilities and availability of computer-aided design and engineering (CAD/CAE). Efficient computational tools developed in this effort can also lead to improved manipulator control algorithms which consider how the effective stiffness, strength, and speed characteristics vary throughout the workspace. For example, no known method exists to analytically distribute the actuators along the manipulator arm with regard to load capacity, spring stiffness, speed of response, etc.

Supporting this effort must be an effective graphics feedback structure to the designer. The mathematical analysis of robotic manipulator mechanisms leads to intricate vector relationships which can not be visualized without graphic assistance. The problem of rendering prismatic, pyramidal, and spherical objects must be solved in real-time as a three dimensional image. Since the functional relationships are known to be highly non-linear, typically dominated by long strings of trigonometric operators, tabular decision rules may be necessary. Specialized hardware may be necessary to calculate the required rotations, translations, and scaling (preferably within nanoseconds). Coloring can be used to provide visual clues for local deformations, actuator load demands, actuator response demands, vibration modes, etc. Given the existence of this technology, it would then be feasible spatially to integrate robotic manipulators in a work cell, to sequentially monitor an assembly process from beginning to end, to study the spatial interaction of two manipulators, etc. to train an operator of a robotic work cell, etc.

7. Sensor Technology

For robotic systems to become more intelligent, they will require a wider spectrum of precision sensors such as proximity, range finding, position, touch, etc. Modern electronic technology has established fabrication techniques at small scales that can be applied to the development of new and/or greatly improved sensing elements. Force sensors deserve special attention because of their irreplaceable role in man-machine communication and enhanced machine intelligence. Integrated strain gauge elements and piezo-resistive films can be deposited directly on compliant structural elements to generate signals to be interpreted by local VLSI electronics. The scale of such devices must match the scale of the task spectrum of the robot. Industrial robots involve force levels of 5 to 150 lbs. and must provide high reliability with minimum compliance. For miniaturized systems, a range of a few ounces and a relatively higher compliance would prove acceptable. Today, exceptionally few such devices are employed, indicating that robotic systems operate at a very low level of intelligence.

Progress has been made on joint position encoders where angular resolution of 20 to 21 bits is now feasible (1 part in a 1,000,000 or 1 arc second) but at high cost ($10,000) and size. Some industrial applications would warrant this resolution when specified end-effector accuracy approaches 0.001 inch. Force sensors of 50 lbs. maximum load and 1 oz. resolution have been developed and are being marketed for approximately $3,000 to $8,000. In themselves, neither of these systems are sufficient to accurately locate the end-effector in world coordinates. The primary dilemma for sensing to-date is that accurate data

on the spatial location of the end-effector (and the rest of the arm as well) without using sensors attached to the arm structure (which introduces deformation errors, noise, and great complexity) remains essentially impossible today. Some acoustic and laser range sensors are known to have resolutions of 1 part in 1000. The resolution of the laser range finder would be enhanced by retro-reflectors on the end-effector. Concepts of optical triangulation, structured light, and laser interferometry have been brought forward to meet this problem. Not only must line of sight be preserved but high speed data reduction would be essential. The lack of this type of technology means that adequate compensations for target deviation is not possible. Hence, robots cannot be taught off-line to do precise operations which implies that the data base of the factory of the future is of nominal value. A solution to this problem would represent a breakthrough in robotic system control and therefore a much broader range of useful applications.

8. Vision

Vision has long been perceived as an important information feedback technology for intelligent machines. With human operators in the control circuit, the use of vidicon cameras is common. These cameras are now available at 2/3 inch diameter with 600 lines of resolution and 5 lux intensity. Solid state cameras are now able to match these properties. These cameras can now display up to 800 by 800 pixels at 2,000 frames per second. Some of the systems can display up to 64 gray levels (only one is known to display color). The primary barrier to the application of solid state cameras for autonomous operation is that scene quantification of visual shape data is very time consuming for the computer system.

Consequently, much of the development effort in recent years has been targeted towards meeting simple scene analysis useful for gross positioning of planar objects or the inspection for salient features such as holes and edges. In the United States, 60 to 80 companies are now offering image analysis systems. Of these, 20 are dealing with complex vision tasks. The most common technique is image or template matching by means of feature extraction (edges, curvature of edges, area moments of inertia, number of holes, array configuration of holes, etc.). One recent offering uses parallel processing and pipeline architecture to treat 350 planar images/sec. with 64 gray levels in "nearly" real time. Another company uses training in terms of a known object using gray scale, textures, color, and light intensity in a combined recognition scheme. Finally, structured light has been used by one firm to check 1250 dimensions on an engine block to a tolerance of 0.015" accuracy over a time period of 35 minutes. The cost of the system could exceed $200,000.

Present vision systems appropriate for integration in robotic systems have a resolution of 1 in 200 or 0.5%, far below that required for tolerance inspection and approximately one order less than the positional resolution of recent precision arms. The number of objects that may be analyzed in the scene is limited by the computational speed of the processor. Further limitations are: required dedicated lighting (preferably as a silhouette), vertical viewing above the planar surface, and limited overlap of the objects. Uniform agreement centers on the need for processing speeds to be increased by two or more orders of magnitude which is probably only feasible by specialized parallel processing architecture or special chips specifically designed for feature enhancement (such as edges). Computational vision becomes more important as the task becomes less structured which implies the need to treat 3-D objects in a generic fashion using color, texture, surface normals, binocular vision, etc. A combination of knowledge base and extremely high speed computation appears to be the only feasible means of achieving real time visual feedback.

9. Artificial Intelligence

Artificial intelligence represents the highest level of decision making required of complex autonomous or semi-autonomous systems such as robots. Machine intelligence implies the implementation of artificial intelligence (decision making) in the operational software of robots. A broad thrust in machine intelligence development is underway in the robotics research community as represented, for example, by several operational robotic manipulator control languages. The overall goal is to create the best structure for the complex layering of performance criteria in coordination with operational means to satisfy some functional task. One facet of this effort is optimal control of multiactuator systems, general path planning, and navigation in terms of imperfectly quantified and statistically changing parameters associated with robotic manipulators.

One of the first steps in creating a complex system is the establishment of a world model which includes the knowledge base, state of the task, expectations based on experience, task times, history of the system degradation, potential obstacles to success, etc. Associated with the world model, is the layered task decomposition of the system operation (i.e., plant—planning, shop—resource allocation, cell—inventory, station—parts handling, machine—part processing procedure, task—motion planning, trajectory—motion control, tolerances—sensor and actuator commands). Optimal task decomposition implies that equal portions of the system's resources are applied to each task level. System operation that is in good agreement with the world model is continued. Otherwise, poor results suggests that this experience should be used to change the world model and the system operation. This may also mean a new task decomposition to better allocate internal resources relative to bottlenecks in the flow of input information (from sensors) or the flow of command information (from processors). As the world model and the balance of task decomposition improves due to experience, the level of internal decision making should diminish and the speed and reliability of this decision making should be enhanced. Research on artificial intelligence should first be pursued in terms of a well defined system having easily measurable performance objectives. Rule-based reasoning can then be used to adapt the artificial intelligence model to the realities of the actual system. This improved world model for AI could then be used to adapt other AI models based on imprecise internal definition and/or fuzzy performance criteria.

10. Intelligent Control

Here intelligent control is intended to mean the global and local control of the system's operation to meet established performance criteria. Status information comes from a series of sensors (tactile, force, visual, etc.) and the data is reduced and interpreted by either distributed or central processors. This interpretation yields command signals to the actuators to carry out the desired operation. One objective is to make the manipulator "electronically rigid" in order to resist all work forces with effectively no deformation and therefore superior precision. Another objective is to make the arm "electronically massless" in order to make system response to commands extremely rapid such that high cyclic speeds can be achieved. A third objective is to make the system parameters "electronically constant" so that system operation, once perfected, would remain invariant. Another representative objective of intelligent control is enhanced smoothness of the prescribed motion in order to reduce shock induced oscillations in the physical structure of the manipulator.

Precision under load is not feasible with today's manipulator technology. In addition to real-time dynamic modeling, a new type of distributed control will become essential in order to provide precision under load. Essentially, the large system

motion is too highly coupled and non-linear to respond to sensory data involving deformations occurring at a much smaller scale. Hence, a new layer of control software and hardware must be developed to treat this small scale function.

Vibration oscillations is the principal limiting factor preventing increased cyclic speeds. Experience with mechanical systems indicates that such oscillations are usually generated from shocks in the command signals. This means that the simplistic start-stop (bang-bang) control of some systems is the worst possible approach. Generalized motion programming synthesized to enhance the smoothness (shocks occur only at the higher derivatives) is now being developed based on the wide experience derived from the programming of read only memory machines such as cams.

Industrial robots do not exhibit perfectly invariant parameters within the complex control and structural subsystems. The sources of the parametric variations may come from changes in actuator electrical resistance (or hydraulic fluid temperatures), friction in joints, dimensional changes due to temperature fluctuations, etc. Implicit parametric variations may also be due to imperfect numerical values used in the deterministic model. The objective would be to characterize these parametric variations and to develop a self-organizing adaptive system to compensate for these variations with reference to the nominal deterministic model. Such a self-calibration system has recently been demonstrated to maintain positional accuracy of an assembly robot.

It is expected that the most advanced level of intelligent control will involve integration of AI principles for perception, feature extraction, image recognition and cognition. This advanced robot control will include unmanned decision making, planning behavior, interaction with other robots, and interface with human operators. AI enhanced control will structure the motion trajectory, timing, avoidance of "obstacles", navigation, strategy, experience driven learning, and lead to a form of anthropomorphic intuition. The actual control of a complex multi-actuator system can only be achieved by voltage commands to and feedback gain adjustment of the actuators themselves. Today, limited success in this final step has been achieved. Intelligent control at this level will be the result of a merger of AI and modern control theory integrating the advances for stochastic and non-linear systems as well as self-organizing and learning adaptive techniques.

11. Software Modules

Clearly software is a critical operational ingredient for robotic manipulators. Thus far a few languages for system control have been developed primarily to enable positional programming and control of the system. As the desired performance of robots is expanded, they will necessarily become more sensor based and more intelligent. But this intelligence will involve an increased level of software. As suggested for actuator modules, the software system will be more rapidly developed and diffused if it is modularized. Then, the system designer will be able to more rapidly assemble a total software package from perfected modules that can be easily debugged or replaced with more effective units when they become available. Such modules could be designed to operate at the highest available sampling rates in hardware dedicated to the software module. Since such modules would be widely used, the associated hardware would become much less expensive.

Appropriate candidates for software modules are associated with smart sensors, prime movers and actuators, end-effectors, and vision. Other modules would depend on the task decomposition of the system. Each task level would involve sensory data from below interpreted by the module combined with commands from above to generate commands to send down and to generate a higher level of information to pass up to the next level. Disturbances or new "world" information could enter horizontally at each task level. At each level several sources of sensory information from below would have to be integrated while the command signal would have to be passed to the

lower level. Thus a generic control structure could be used at each level. The format, sampling rate, and quality of information would be dependent on the height in the hierarchical tree. The total computation effort per hierarchical module would be kept as constant as possible to reach optimum results.

12. Computer Architecture

The growth of the field of micro-electronics is the primary reason that accelerated development in robotic systems technology becomes feasible today. The goal is the distribution of micro-electronic hardware modules throughout the system (distributed intelligence) in order to make smart sensors, smart actuators, etc. Because these hardware modules are each dedicated to a unique task, the calculations could be made more rapidly in special high speed arithmetic processors. Another opportunity is the trade-off between computational speed and precision for robotic motion. The model is based on CORDIC perhaps best known for its use in hand-held calculators. Basically, the process is an iterative procedure for function evaluation (such as trigonometric functions) where with each iteration, additional precision is obtained. Hence, speed and accuracy are natural trade-offs in this process. In robotic systems, precision is usually required at only a few positions in the cycle. The required equipment would involve only standard firm ware components, notably bit-slice processors and high-speed look-up tables (ROMS). For example, the TRW 24 x 24 bit multiplier chip can be used to multiply two 32-bit floating point numbers in 200 nanoseconds. This speed is 400 times faster than the fastest multiply statement executed on the DEC PDP 11/23.

One of the primary problems limiting progress towards real-time operation of intelligent robots is that existing serial processors are poorly suited to treat the fundamentally parallel nature of the phenomena of robotic manipulators. For example, future systems may involve many sensors generating a large information array all of roughly equal significance to the system. This reality of excess data, all at the same level, has been adequately demonstrated for machine vision. It is much less well understood with regard to the real time operation of the dynamic model of the manipulator system. There are six distinct computational levels which must be implemented serially. Within each of these levels 100 to 800 distinct independent functions can be calculated in parallel. This massive functional parallelism shows that parallel processing is essential for the real time control of any system having the geometric complexity of a general robotic manipulator. As mentioned for spatial end-effector position sensing, an economical parallel processing architecture would represent a break-through for the next generation of robotic system.

13. Communication Interfaces

Many practitioners in robotic implementation have discovered communication mis-matches between system components (primarily at the machine level). However, as the data base of the factory of the future becomes more addressable, the need for very highly integrated communications will become imperative. Since no one manufacturer will supply all factory units, standardized interfaces will become very desirable. At the other end of the spectrum are the interface needs between robotic components such as sensors, actuators, distributed processors, etc. Some of the issues are voltage levels, rates of sampling, numbers of channels, multiplexing, AD-DA converter technology, scaling, synchronization, error filtering, noise reduction and isolation, and data compaction. Obviously, both hardware and software issues are involved. The goal must be to standardize as many of these interfaces as possible. The National Bureau of Standards robotics program is pursuing this objective as one of their major missions. The Navy is working on ways to establish accurate long range communication with untethered vehicles in the difficult medium represented by ocean sea water which contains debris. In the Oak Ridge National Laboratory fuel reprocessing plant development, tethers would drastically limit mobility of the maintenance and handling equipment. Hence, special frequency radio wave systems are planned to ensure complete mobility.

309

14. Man-Machine Interface

Almost all of the development work now being pursued in the U.S. deals with autonomous machines. This approach assumes that artificial intelligence can be transformed into an operational machine intelligence capable of duplicating or exceeding the judgement and decision making capability of the human operator. For repetitive, and highly structured tasks as occur in simple manufacturing processes (pick-and-place, spot welding, spray painting, etc.) this is possible. For tasks such as complex assemblies, nuclear reactor maintenance, or avoidance of maneuvers of an intelligent enermy, the required level of intelligence does not appear to be feasible in the next two decades. The best near term opportunity is to use a balance of human and machine capabilities. As the machine technology improves, less will be asked of the human and more of the machine. This man-machine approach allows the most rapid penetration of the manufacturing market with near-term technology, allows a gradual and natural transference to more machine-oriented systems, and allows a minimum disruption of the manufacturing workforce.

The objective is to develop a transparent and universal interface between the human operator and the robotic manipulator. Commands to the manipulator from the human must be made in the most natural manner (voice, digital, or kinesthetic) and must occur with a minimum burden on the operator. Force feedback is critical to the full awareness of the operator. In other words, the interface must be optimized for the most effective use by the operator. Also, information derived from sensors on the slave manipulator must be enhanced by the interface software in order to make it as useful as possible to the operator. Since the human's instincts are to operate in real-time, the sampling rate of the system must exceed 30 hz. The manual controller is effectively a light weight robot which drives the slave manipulator through digital commands. Hence, this system is essentially equivalent to two cooperating robots. This is why numerical interfacing of a manual controller and a robot manipulator is much more difficult than operating an autonomous robot.

III. CRITERIA FOR ADVANCED ROBOTICS TECHNOLOGY

The following is a listing of 14 distinct criteria that may be used as indicators of the level of the technology available in an advanced robotic system and may be a useful means to judge progress of the technology under development.

1. Multi-task capability means the number of different physical tasks that can be performed by the same robotic system.
2. Level of machine intelligence implies the level of integration of computer hardware, software, and artificial intelligence to make the system as autonomous as possible.
3. Time efficient operation implies the speed at which the robotic system performs its task relative to the human performing the task alone.
4. Unstructured task level suggests the level of numerical uncertainty of the operation that is to be performed by the robotic system.
5. Geometrical dexterity is an indicator of the motion range the end-effector can move through while performing physical tasks.
6. Portability and mobility implies the level of movement the total robotic system has relative to a stationary (fixed shoulder) manipulator.
7. Precision is an indication of the absolute precision of placement of the end-effector in world coordinates in response to simple numerical commands.
8. Load capacity clearly implies the ability of a robot to carry or resist a given load without major deformation.
9. Reliability is an indicator of the failure rate of the total robotic system.
10. Obstacle avoidance suggests the ability of the robot to avoid obstacles in its work environment.
11. Force sensing suggests the measurement of forces in the manipulator system to be evaluated by the machine intelligence to judge working forces or to compensate for manipulator deflections.
12. Smoothness of operation implies the lack of backlash or very large deformations in the manipulator system.
13. Operational envelope gives an indication of the working range available by the robot without moving its shoulder.
14. Vision corresponds to shape recognition either by analog feedback to the human operator or by digitizing the scene and providing numerical shape recognition.

511

TABLE 5: ESTIMATE OF LONG TERM IMPORTANCE OF ROBOTIC CHARACTERISTICS FOR VARIOUS APPLICATIONS

Estimate of Importance of Robotic Characteristic for Various Applications / Robotic Characteristic	Average	Micro-Processing	Assembly	Light Machining	Welding	Riveting	Fission Reactors	Fusion Reactors	Oil Production	Coal Production	Fuel Handling	Ocean Operations	Battlefield Op.	Emergency Repair	Ammunition Hdlg.	Strategy Develop	Micro-Surgery	Prosthetics	Agriculture	Accident Missions	Service Robots
1. Multiple Task Capability (2)	6.3	4	5	3	8	4	10	10	10	5	5	8	8	7	3	5	8	4	4	8	7
2. Level of Machine Intelligence (1)	7.1	5	7	8	8	10	7	7	7	3	5	9	9	6	5	10	7	4	4	6	5
3. Time Efficient Op. (6)	5.4	6	6	4	5	5	7	7	6	3	3	7	6	5	6	7	6	3	8	5	3
4. Unstructured Task Level (7)	5.2	2	2	2	7	4	7	7	7	4	3	7	7	7	2	7	7	4	4	10	5
5. Geometric Dexterity (10)	4.8	4	6	4	7	3	7	7	7	4	3	7	6	5	4	-	7	3	4	6	3
6. Portability & Mobility (4)	6.1	2	2	2	6	6	7	7	10	10	5	10	8	6	4	-	3	8	8	10	8
7. Precision (5)	5.5	10	7	10	7	7	5	6	6	5	4	5	5	6	6	-	10	3	3	3	3
8. Load Capacity (12)	4.7	1	3	10	7	7	5	7	6	6	4	5	5	4	8	-	2	3	4	4	3
9. Reliability (3)	6.3	5	5	7	5	5	5	6	7	6	8	7	7	5	8	7	6	7	7	8	5
10. Obstacle Avoidance (8)	5.1	4	10	5	8	3	4	4	4	5	5	7	5	5	7	-	4	3	5	7	7
11. Force Sensing (11)	4.8	7	7	10	6	6	3	4	4	4	4	5	5	5	3	-	10	4	3	3	3
12. Smoothness of Operation (13)	3.8	7	5	10	5	3	2	2	2	2	2	3	3	3	3	-	10	5	3	3	3
13. Operational Envelope (14)	3.1	4	7	3	6	5	2	3	2	3	2	2	2	2	4	-	4	3	3	3	3
14. Vision (9)	5.0	10	10	3	8	7	2	3	2	4	5	7	6	5	2	-	8	2	8	4	4

TABLE 6: IMPORTANCE OF ROBOTIC CHARACTERISTICS BY APPLICATION GROUPS

Robotic Characteristics	Average for all Applications	Industrial Automation	Energy Systems	Military Operations	Human Augmentation and Agriculture
1. Multiple Task Capability	6.3	4.8	8.0	6.2	6.2
2. Level of Machine Intelligence	7.1	7.6	5.8	7.8	5.2
3. Time Efficient Operation	5.4	5.2	5.2	6.2	5.0
4. Unstructured Task Level	5.2	3.4	5.6	6.0	6.0
5. Geometric Dexterity	4.8	4.8	5.6	5.5	4.6
6. Portability and Mobility	6.1	3.6	6.8	7.0	7.4
7. Precision	5.5	8.2	5.2	5.5	4.4
8. Load Capacity	4.7	5.6	5.6	5.5	3.2
9. Reliability	6.3	5.4	6.4	6.8	6.6
10. Obstacle Avoidance	5.1	6.0	4.4	6.0	5.2
11. Force Sensing	4.8	6.2	3.8	4.5	4.6
12. Smoothness of Operation	3.8	6.0	2.0	3.0	4.8
13. Operational Envelope	3.1	5.0	2.4	2.5	3.2
14. Vision	5.0	6.6	3.2	5.0	5.2

For all applications, the most important robotic characteristic does not outrank the least by more than a factor of two. The range is up to a factor of 4 among some of the application groups. This data is partially tabulated below in order to establish the most siginificant properties of robotic systems for each application group. Generally, as the application warrants or allows autonomous operation, the characteristics of machine intelligence, precision, vision, sensing, and reliability become important. For unstructured task applications requiring a balance between man and machine, characteristics such as multiple task capability, mobility and portability, obstacle avoidance, reliability and unstructured task level have an increased importance.

		Component	Rank
All Group Applications	—	Level of machine intelligence	10.0
		Multiple task capability	9.0
		Reliability	9.0
		Mobility and portability	8.6
		Precision	7.8
		Time efficient operation	7.6
Industrial Automation	—	Precision	10.0
		Level of machine intelligence	9.3
		Vision	8.3
		Force sensing	7.5
		Smoothness of operation	7.3
		Obstacle avoidance	7.3
Energy Systems	—	Multiple task capability	10.0
		Portability and mobility	8.5
		Reliability	8.0
		Level of machine intelligence	7.2
		Load capacity	7.0
		Geometric dexterity	7.0
		Unstructured task level	7.0
		Time efficient operation	6.5
		Precision	6.5
Military Operations	—	Level of machine intelligence	10.0
		Portability and mobility	8.9
		Reliability	8.7
		Time efficient operation	7.9
		Multiple task capability	7.9
		Obstacle avoidance	7.7
		Unstructured task level	7.7
Human Augmentation & Agriculture	—	Portability and mobility	10.0
		Reliability	8.9
		Multiple task capability	8.4
		Unstructured task level	8.0
		Vision	7.0
		Obstacle avoidance	7.0
		Level of machine intelligence	7.0

One of the principal responsibilities of a research team is to develop the technological criteria necessary to measure the impact of proposed or actual advances in that technology. For robotics, as an immature field, many of the criteria are new and relatively unknown in their overall importance to the resulting system's operation. The following 14 factors should prove adequate to define and evaluate a generic robotics technology.

1. Multi-Task Capability

The operational task spectrum of most industrial robots is severely limited. Some are limited to a single function such as pick-and-place. Others can perform sequential spot welds or pre-programmed painting. The most advanced system of this type can perform approximately 20 distinct operational functions.

The concept of multi-task capability means that a wide range of functional tasks can be performed by the same robotic system. This concept can be illustrated by the example of PWR steam generator maintenance where the sleeving task may require up to 25 sequential sub-tasks all representing distinct operational requirements. The steam generator presently requires 18 tasks such as plugging, sleeving, etc. The nuclear steam system of PWR's represents 10 distinct system component tasks such as the steam generator, pumps, valves, etc. The combined generality of system tasks, component tasks, and sub-tasks is the primary reason why a generic technology is essential for a multi-purpose robotic system operating within an unstructured environment. Should the unstructured nature of the task be articulated by unknown or unfriendly forces, the need for generic technology becomes even more critical.

2. Level of Machine Intelligence

The primary objective of machine intelligence is to produce a quality of motion at least equivalent to the human acting alone. To accomplish this level of performance requires a level of sensibility to the operational environment and supporting intelligence similar to the sensing and reflex action (distributed intelligence) in the human arm. Since no robotic system today exhibits any significant level of intelligence, effective integration of machine intelligence would provide a real opportunity for improved performance. Simple off line pre-programming is insufficient to treat the unstructured task spectrum described in item 4 below. A combination of human intelligence and machine intelligence in a balance best suited to perform a given range of tasks is recommended for the performance of all but the simplest structured tasks.

3. Time Efficient Operation

In many robotic functions, the time required to perform a given task may have significant economic impact or it may be crucial to the overall effectiveness of the task being performed.

For nuclear reactor maintenance, the availability of the reactor for power production is a major economic issue. In military operations, time may be essential in response to a surprise attack or a rapid change in tactical plans. The essential requirement is that the robotic system be at least man-equivalent in this regard. In nuclear reactor maintenance, the goal should be to reduce task times by 50% which would have hundreds of millions of dollars per year benefit. The present

light duty hot-cell master/slave systems exhibit task performance eight times slower than the human acting alone. However, if the human operator is removed completely from the hazardous environment, other benefits accrue since the environment can be greatly simplified (reduced cost) or more rapid start-up can be achieved. For this reason, an improvement of two or three times in direct task time performance (over the present technology) by the robotic system may well prove sufficient to achieve overall task times one-half of those for the human acting alone.

4. Unstructured Tasks Level

Here the concept of an unstructured task means that the operational environment is not quantitatively known to the operator, to the machine intelligence, or to the data base. Many systems such as nuclear reactors are documented as designed not "as built" and they frequently are not provided with any reference benchmarks. This means that sensing feedback (both force and visual) is essential to the performance of unstructured tasks. Machine intelligence enhances this perception and makes system performance more accurate and rapid. Generally, most existing systems for remote operations provide a modest capability to treat the lack of definition represented by the unstructured task.

5. Geometrical Dexterity

Geometrical dexterity is meant here to denote end effector motion of great generality in space. The human hand moves with a first level of dexterity augmented by the additional 6 DOF supplied by the human shoulder. Using a fixed shoulder would dramatically limit the human arm's dexterity. The ability to analyze arm geometry is now well established. To design for a required level of dexterity has been shown to be feasible and progress is being made. One of the best ways to increase dexterity is to add 2 DOF to make an 8 DOF arm. These extra (redundant) DOF makes obstacle avoidance much more likely. Unfortunately, these redundant DOF make the control of such an arm very difficult. A solution to the dexterity design problem is required in order to provide the designer an essential tool to select the best possible manipulator geometry.

6. Portability and Mobility

Portability of the robotic system implies that it can be broken down into modules which can be carried to the work place by a human operator and quickly assembled. The suggested weight limit per module is 35 lbs. Such a weight restriction creates an unusual demand to design light weight actuators and to use special light weight materials (composites or carbon fiber).

Mobility implies that the system could move over (or traverse) an obstacle strewn area. To date, no such system exists in the general sense. Special tracked vehicles, track followers, and wheeled vehicles are used to traverse relatively smooth surfaces (or fixed tracks) with minimal obstacles. Unfortunately, for many applications, these special conditions do not exist. Mobility would have special significance to surveillance and to dedicated autonomous units for military applications, accident missions, and ocean floor activity. During the past 20 years, significant laboratory work has been on-going on the generic concept of walking machines for mobility purposes.

7. Precision

The absolute precision of most industrial robots is known to be not better than 0.05 inch and many are far less accurate. Yet, many assembly, welding, and light machining operations require a precision of 0.01 inch. Further, fine positioning to 0.001 inch is sometimes necessary. For the example of nuclear reactor maintenance, the overall need, with regard to precision, is equivalent to that of a portable machine shop. This level of precision puts an unusually demanding resolution requirement on the actuators and their control system. The control encoders and actuators must be capable of steps of 10 seconds of angular rotation. Most actuators fall far short of this, especially if they must provide a high load capacity. In addition to these precision requirements, the more difficult condition is to maintain precision while the manipulator experiences large load variations. It is common for external loads to degrade the unloaded precision by a factor of ten. The reader can prove this reality to himself by "shaking hands" with a few industrial robots. It is not uncommon to easily achieve oscillations of 1/4 inch in magnitude.

8. Load Capacity

The load capacity of the arm is primarily dependent on the size of the arm's actuators. Generally, about 90% of the arm's deformation occurs at the actuators. Today light duty arms are designed to carry 10 lbs. Infrequently, arms are designed to carry 200 lbs. but they are heavy, imprecise, sluggish, and certainly not portable. A load capacity of 200 lbs. is recommended for stream generator maintenance in nuclear reactors. In micro-surgery, load capacity may be measured in ounces. One of the best ways to improve load capacity is to place the actuators in a parallel structure so that they can be carried by the base and not by the arm as they are presently for serial manipulators. Another useful effort is to seek an optimal distribution of actuator sizes in a given arm geometry.

9. Reliability

Industrial robots, today, have established a very high operating availability of approximately 98%. These units are marketed only after prolonged testing and redesign. Nonetheless, in other unique applications, this extensive history is not available to ensure high realiability. This property is especially important in such operations as nuclear reactor maintenance. Failure would mean difficult retrieval and an extended down time (at great cost) of the power plant. Here, the goal is failure in 1 of 20 field operations (each lasting 2 to 5 days). Failure is also unacceptable where human life is involved as in accident missions, military operations or ocean floor activity. Predictably, the simpler systems having lower intelligence will be substantially more reliable. Hence, it can be recommended that for an integrated system with all technologies implemented, numerous field demonstrations will be necessary to perfect the system in order to make it sufficiently reliable.

10. Obstacle Avoidance

Many unstructured tasks must be performed within a volume containing known, unknown, or moving obstacles. Today, working in an obstacle strewn environment is rarely considered in automation operations on the factory floor. Almost no existing robot has a significant level of obstacle avoidance capability although those which are anthropomorphic are more able to avoid obstacles. In the case of steam

generator maintenance, access is difficult. In nuclear reactor systems, the maintenance operation for piping and valves is heavily obstructed by obstacles. The best way to achieve increased avoidance capability is to increase the generality of the arm's geometry. Beyond this, increasing the DOF to 8 will prove very beneficial. Unfortunately, both of these steps make the design and control problems much more difficult. Having the increased generality makes increased machine intelligence essential to benefit from proximity sensors on the arm.

11. Force Sensing

Force sensing is the most basic sensing parameter necessary for feedback to the operator or to the machine intelligence of a robot operating in an unstructured task regime. Other sensors are tactile sensors in the fingers of the end-effector, torque sensors at the actuators, etc. An accurate level of sensing should dramatically improve the system's operation making it possible to perform such functions as hammering which are essentially impossible today. Assembly operations are known to be significantly faster and more reliable with force sensing in the system. Some servo master/slaves exhibit a reasonable level of force sensing today. Unfortunately, the master/slave system can not easily be generalized and does not lend itself easily to the integration of machine intelligence.

12. Smoothness of Operation

Smoothness of operation of the system implies that no unexpected or unpredictable phenomena disturb the human operator or the machine intelligence in the performance of the operation task. These disturbing phenomena are backlash and large system deformations. Present light duty arms avoid backlash but they exhibit very high deformation under load. Present heavy duty arms may allow as much as 1/4 inch backlash at the end-effector. Advanced system design must avoid these pitfalls.

13. Operational Envelope

The reach of the arm directly affects the size of the operational envelope or field of movement of the manipulator arm. Small arms (of 3 ft. reach) tend not to be able to duplicate the scale of human motions. Many maintenance tasks for nuclear reactors and some military applications require arms of 6 ft. in length. Unfortunately, the stiffness of these arms is inversely proportional to the cube of its length; i.e., it becomes compliant very rapidly. But the reach concept of the arm is much more involved than it first appears. To be able to approach an extreme position and remain dexterous is usually not possible. As one approaches the limits of the operational volume, dexterity deteriorates rapidly. Maintenance tasks such as steam generator sleeving require high dexterity throughout the work volume.

14. Vision

The sensing information by analog or computational vision is known to be an essential ingredient in the operation of robotic systems in unstructured task regimes. This information may go directly to the human operator or to the machine intelligence, or to both. Recent progress in analog vision has been sluggish and no breakthroughs are expected. Analog vision displayed for the human operator enhanced by machine intelligence is an untapped opportunity. Furthermore, the use of digital vision or graphics could be valuable in training systems for the location of obstacles and other features. Vision technology can be enhanced by better integration of automatic camera control and foveal vision.

III. SUMMARY REPORT ON PIPELINED COMPUTATION OF DYNAMIC MODELING MATRICES FOR SERIAL ROBOTIC MANIPULATORS

BY

JOHN P. WANDER

AND

DELBERT TESAR

As robotic manipulators are called upon to perform a greater variety of tasks, increased dynamic performance is required forcing the development of improved control algorithms. Such algorithms are invariably tending to require a reference model of the manipulator to be employed in an adaptive control scheme. To date, the computation of such a model has been considered a time consuming task even for off-line simulation purposes. The application of array processing techniques to this problem is facilitated by a decoupled form of the controlling equations. A decoupling of velocity and acceleration from position dependent terms results if the equations for the inertia torque of an N-link serial manipulator are written as:

$$\underline{\bar{T}} = [I^*]\underline{\ddot{\theta}} + \{\underline{\dot{\theta}}'[P_1^*]\underline{\dot{\theta}},\ \underline{\dot{\theta}}'[P_2^*]\underline{\dot{\theta}} \ldots\ \underline{\dot{\theta}}'[P_n^*]\underline{\dot{\theta}}' \ldots \underline{\dot{\theta}}'[P_N^*]\underline{\dot{\theta}}\}'$$

where:

N is the number of links

$[I^*]$ is the generalized inertia tensor, independent of $\underline{\dot{\theta}}$ and $\underline{\ddot{\theta}}$

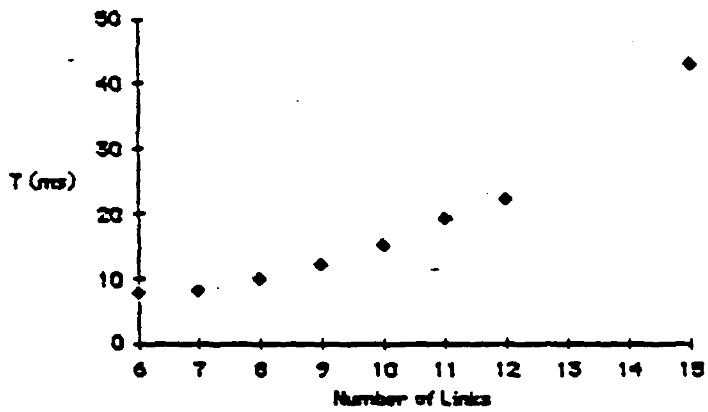
$[P_n^*]$ is used to compute the torque at joint n due to the change in $[I^*]$.

It is an important property of the matrices $[I^*]$ and $[P_n^*]$ that they are functions of the generalized coordinates only and not their time derivatives. These matrices represent the model of the manipulator required for improved manipulator control.

Since the matrices are functions of the generalized coordinates only and their expressions involve repeated calculation of a fixed algorithm, the modeling is able to be efficiently computed with pipelined code. Because there are relatively many (in the context of pipelined computation) small vector and matrix operations requiring predictable but complicated addressing of intermediate results, the pipeline must be able to address operand data via precomputed offsets also available as data. Thus the algorithm development for computing the modeling matrices involved two distinct tasks - development of the software to compute the correct offset vectors for arbitrary model size and development of the pipelined algorithms for the real-time computation.

The resulting software is written in FORTRAN and pipeline code for the Analogic AP-500 consisting primarily of Motorola 68000 macros. The

offline portion runs in the host computer and then loads the AP-500 with the offset data. The offline software is initiated from an interactive environment from which the user is queried for a description of the manipulator. The real-time modeling is designed to be triggered by the transfer of the generalized coordinate vector into the AP. For a completely general, serial, six-link manipulator the modeling requires 7.5 ms which yields an update frequency of better than 100 Hz. The elapsed time results for several different sized models are shown in the Figure. The computation time required by the modeling algorithm grows in proportion with the fourth power of N, the number of links in the manipulator modeled.



Actual Computation Time For Serial Manipulator Model

The conclusion that can be drawn from these results is that evolving computer technology will make extensive dynamic modeling information available in real-time for use in improved manipulator control schemes. Here the use of an array processor instead of a mini computer has produced a cost/benefit ration of 50 to 1. Further advances in pipelined architectures, both in hardware and in software, as well as the implementation of distributed processing techniques will undoubtedly allow the control algorithms for manipulators to update gains based on the modeled dynamic system. It will not be long (perhaps 5 years) until models of flexible manipulator dynamics run fast enough to allow control of lower vibrational modes of the manipulator structure.

IV. AN ASSESSMENT OF THE DEVELOPMENT AND APPLICATION POTENTIAL FOR ROBOTS TO SUPPORT SPACE STATION OPERATIONS

BY

DELBERT TESAR

ABSTRACT

NASA is developing a space station platform to provide satellite servicing, laboratory operations for micro-g gravity experiments and manufacturing, and support for military space activity including perhaps the strategic defense initiative (SDI). This development will require two decades and a total investment of $8 billion (plus) of which 8 to 13% (or more than $1 billion) will be set aside for automation and robotics. The need for robotics stems from the hazardous nature long-term EVA operations present to astronauts and the fact that too many such functions would overburden crew time or be beyond the physical capability of the astronaut. Extensive analysis of the maintenance operations required of robots for the space station indicated that they are essential to make the station economical and perhaps even feasible.

The first reality for the space station is that it is a one-of-a-kind effort and that its operation will likely parallel the experience gained in nuclear reactors which are available 75% of the time. This lack of availability in the space station would be devastating to its usefulness to the military and especially to the strategic defense initiative. The best means to assure an availability approaching 95% would be through the full implementation of an advanced and balanced robotics technology to perform service and maintenance especially under emergencies or attack.

Five principal applications for robotics for the space station are:

Assembly of space structures
Space station maintenance and repair
Satellite servicing and repair
Hazardous manufacturing and laboratory experiments
Maintenance of robots

of which 80% of the activity will be associated with satellite servicing. These operations involve a range of physical tasks which will require that the overall technology base for robotics be moved significantly forward from the second generation industrial robot of today to the fourth generation adaptive and modular robot feasible within 10 years if a major R&D program is established now.

The dominant aspect of the technology for robotics needed for the space station is that it must be light weight and highly versatile, capable of performing a very broad range of physical tasks, some of which may require precision under disturbances. This forces attention on a balance of component technologies (14) of which the top 7 long-term priorities are:

Man-machine interface
End-effectors
Actuator modules
Sensor technology
Computer architecture
Graphics/CAD
Intelligent control

Similarly, the progress of a major national robotics program must be measured in terms of a finite number (14) of system criteria for the operation of

robotics in the space station. The top 7 long term criteria are:

- Multiple task capability
- Level of machine intelligence
- Precision
- Portability and mobility
- Reliability
- Obstacle avoidance
- Force sensing

The conclusion derived from a matrix analysis for these priorities in this report is that no one component technology or system criteria can solve the mission needs of space station robotics. What is critically needed is a balanced development of all component technologies and system criteria in proportion to the demonstrated needs derived from a careful functional analysis.

If these needs are met for the space station, much of the technology required for the next generation of industrial robotics will be made available which should result in a full implementation of robotics in the factory of the future. The tentative cost summary for the automation and robotics technology development suggests that 56% be allocated to man/machine interface and robotics and 44% be set aside for information management with 82% for research and 18% for prototype demonstrations. These breakdowns appear to be sound although somewhat more could be warranted for demonstrations of the robot system prototypes.

TABLE OF CONTENTS

I. INTRODUCTION

NASA now has the task, defined and supported by Congress and the President, to establish a station in space to provide the next level of technology and function beyond that provided by the shuttle. The proposed system (see Fig. 1) may act as a fuel depot for other space systems, to provide a base for assembly and repair for those systems, and to carry out special earth monitoring and space manufacturing development unique to a stable platform in micro-g space. Extensive review* by several industrial firms and independent analysis groups shows that the overall technology base, particularly in space station automation and robotics, must be significantly improved in order to reduce the cost of the station, and perhaps, make it economically feasible. Secondary objectives are to make the space station more efficient and to enhance the tech base of U.S. manufacturing. These objectives are being pursued under a budgetary allotment of 8 to 13% of the total space station program for the purpose of automation and robotics. An SRI study report recommended that in excess of $100 million/year be set aside to pursue automation and robotics for the space station.

The environment for humans in space is far from friendly (high vacuum, extreme glare, extremes in temperature, occasional high levels of radiation, etc.), requiring the astronaut to prebreathe oxygen for an hour before EVA to avoid the "bends". Space suits limit the dexterity of astronauts making delicate operations (such as satellite repairs) or space structure assembly (of very large modules) very time consuming and taxing. The prospect of performing hazardous experiments and manufacturing in the space station raises the question of safety for the crew, especially during failures or maintenance. Finally, the power plant for the station may very well be nuclear, whose radiation would further restrict the movement of astronauts. For all these reasons, it is now perceived that robotics is critical to the deployment and operation of the space station.

In carrying out this mission, NASA has also been asked to make the technology useful in enhancing U.S. productivity. In order to meet this objective, NASA must evaluate the weaknesses plaguing the nation. The economic reality is that wealth generation is 1/3 of the GNP of the U.S.A. of which 2/3 is due to manufacturing. Of this 2/3, 60% or more is mechanical in nature (shoes, clothing, cameras, industrial machine tools, etc.) primarily in the civil sector. Yet industry invests less than 6% of its R&D and its manpower to meet this need. The federal government does even less, investing no more than 0.7%. These major imbalances between input R&D and output economic reality continue to weaken our civil sector and allow penetration into our home markets by other strong civil sectors, i.e., Japan, Germany, and in the future, France. Hence, if NASA is going to provide the tech base to reduce the effects of these imbalances, NASA will have to purposely develop a balance of technologies in the automation and robotics program. Generally, it appears that NASA is moving cautiously in this direction by expressing concern about too much reliance on unproven technologies, by expressing a very high regard for the man-machine interface, and by laying out some elements of a new mechanical technology for robotics.

* See Ref. 1-9.

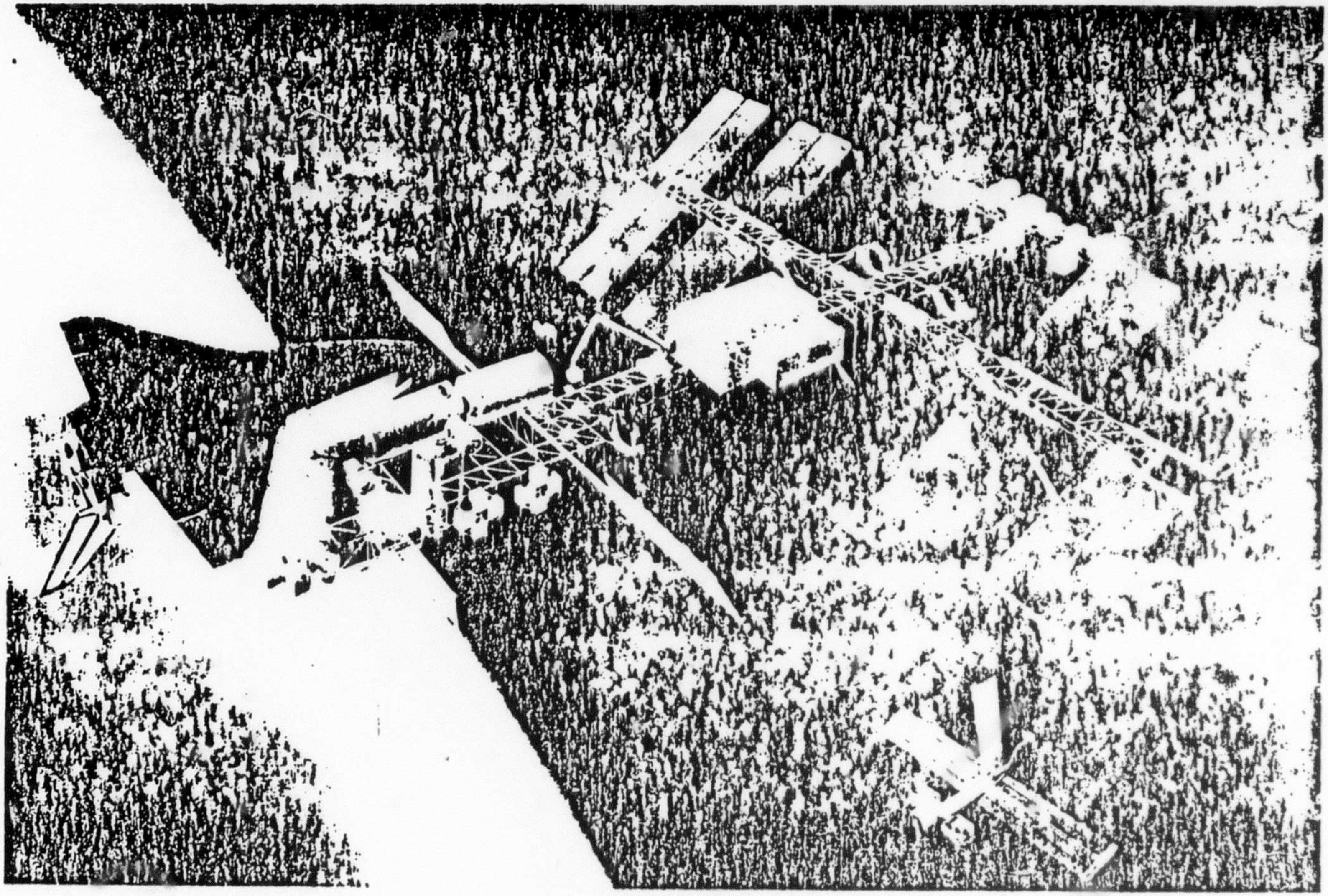


FIG. 1 CONCEPT OF NASA SPACE STATION

Since very few astronauts can be housed in the station, they must be made as effective as possible, making the day-to-day housekeeping of the station a prime target for machine intelligence. Also, the astronaut must be augmented as much as possible, making expert systems a truly valuable aid, especially in unusual or unexpected circumstances. Beyond this, it appears that extending the manual dexterity of the astronaut through "telepresence" becomes essential for repair, maintenance, and remote emergencies outside the protected envelope of the station cabin.

This report is written with this last objective in mind. The question to be dealt with is:

> <u>What research priorities must be established now to create a balanced robotics technology to meet the special needs that exist in space and that would prove helpful in enhancing U.S. productivity in manufacturing?</u>

As a consequence, this report will deal with technologies of the next generation (and beyond) robot with a view to the effective balance between electrical and mechanical component technologies.

II. GENERAL TECHNOLOGY BASE ISSUES FOR ROBOTICS IN SPACE

Experience derived from the maintenance of nuclear reactors shows that 40% of the downtime is due to forced outages (the unexpected). Similarly, it can be expected that a large portion of the need for robotics in space will be to deal with unique and widely varying maintenance operations. This broad capability to react is difficult to imbed in a few astronauts (even in the unlikely event that they will have sufficient time to give to the emergency). Hence, every assistance to the astronaut must be provided to make the space station an operable system. This includes expert systems, man-machine interface, and dexterous multipurpose robots. Paradoxically, as the intelligence level in robotics improves, there will be a greater need for a transparent interface between man and machine, not less.

The dominant aspect of the technology for robotics for the space station is that it must be lightweight and highly versatile, capable of performing a very broad range of physical tasks, some of which may require precision under disturbances. The space station is likely to be highly flexible and susceptible to oscillations if impacted by docking forces. Such oscillations would probably prove disastrous to sensitive manufacturing processes or to delicate experiments specifically planned for the space station because of its high level of stability. Hence, this means that a whole new technology of "soft docking" using carefully controlled robotic berthing structures will become essential. This type of dynamic adaptive control does not exist in the literature in any form, but it does represent a major opportunity to move the technology forward (this might be called adaptive dynamic control).

Another physical disturbance to the station could occur due to the "walking" of a multi-legged robot over the structure. This problem, which can easily be ignored on earth, is probably the most difficult (but interesting) problem facing the mechanical engineer. It means that the forces on the "feet" of such a robot must be carefully balanced at all times to reduce the disturbance to the structure. This balance means that up to 24 inputs must be in equilibrium dynamically at all times (in real time) to create the desired

output motion (construction of the space structure itself) while not stressing, bending, crimping, oscillating, etc., that structure.

Smaller robots will be essential for the operation of the manufacturing facility or the experimental research facility. These robots must be lightweight, modularized for easy repair, highly dexterous to perform delicate operations, capable of doing precision light machining, and redundant (extra DOF) to allow obstacle avoidance. These are very demanding criteria, far beyond the technology represented by today's industrial robot which is heavy, has a low level of dexterity, can perform only those functions which generate no disturbances, has a very limited integration of modern control technology, is programmed in the most archaic manner, and has generally a very poor man-machine interface.

The lightweight requirement means that the arms are going to be made of flexible links which will yield large deformations under load. This type of system can be made "electronically rigid" only by the most sophisticated of dynamic modeling and real time control. Not doing so will mean that large deformations will either confuse the human operator (beyond his ability to correct) or will make the database in the computer essentially useless.

In the micro-g environment, the weight of the arm (and the weight of its load) is not a control or disturbance issue but its mass generated inertia loads are. Hence, the dynamic programming of the robot to minimize shock (and the resulting oscillations) to the system becomes essential (a topic long understood for precision ROM machines such as cams). Dynamic shock can also occur from built-in backlash with the worst possible results when working in the micro-g environment. All this means that a full parametric model of the arm must be developed, computed in real time, and controlled by sophisticated adaptive methods only now being conceptualized in research programs. It must be remembered that the robot is an N DOF highly non-linear dynamic structure from which precision output is desired—a problem beyond the accomplishment of the control field at this time. The space robot compounds this complexity by being "limber", redundant, modular, and accessible to direct control by both man and computer.

Finally, several analyses of space station maintenance needs shows that at least two robots (see Fig. 2) will have to cooperate with each other to carry out dexterous tasks because gravity does not hold the work piece in place as it does on earth. These cooperating robots will have not less than 12 inputs combined which must act to create no more than 7 outputs (screwing a nut on a bolt). This force balancing (in real time) among 12 inputs to create 7 precise output forces is beyond today's control technology. Nonetheless, it is probably one of the most interesting control problems that can be posed because it forces a full integration of computer science, artificial intelligence, mechanical system modeling, modern control theory, and adaptive control. Some of the operations by these cooperating robots will be precision light machining (drilling, routing, trimming, etc.) which will cause disturbances to the robot structures which must also be dealt with.

Hence, the opportunity for robotics offered by the space station mission is truly of the greatest possible importance and, if carefully managed, could be instrumental in developing the next generation of technology (and beyond) for robotics. As mentioned before, progress will be made only if balance in the component technologies is the guideline in structuring the program (see Sec. VIII). Obviously, if these needs are met for the space station, a full implementation of robotics in the factory of the future would be assured.

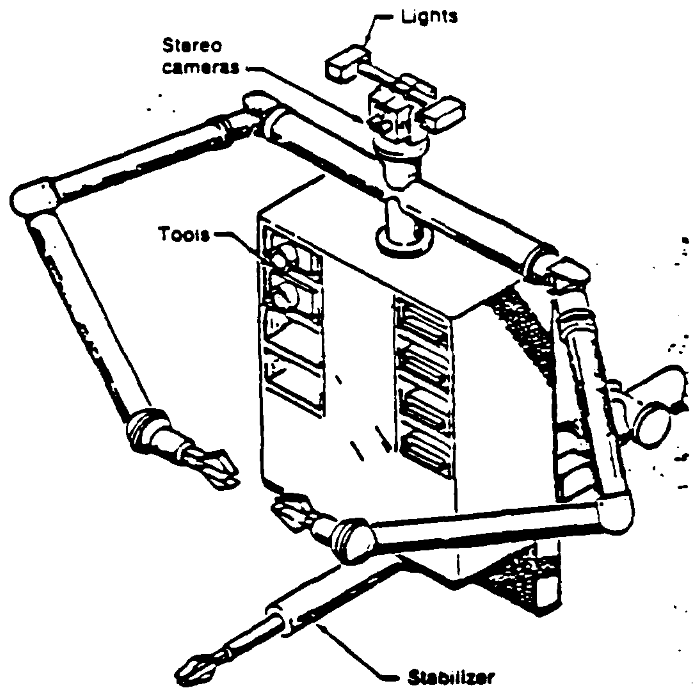


FIG. 2 DUAL ARM MANIPULATOR REMOTE SERVICING MODULE CONCEPT (MARTIN MARIETTA, 1984)

III. ASSEMBLY OF SPACE STRUCTURES

One of the first functions that must be planned for is the assembly of the space station. This involves a wide spectrum of tasks from the assembly of large modules to small mating tasks such as bolting, locking, and forming joints in the structure itself. This assembly process could be highly structured minimizing the level of uncertainty. A pallet with structural elements would be positioned near the assembly system (conceptualized in Fig. 3) which could contain two or more arms, some for gross motion and at least one for fine motion.

In the IOC (1991) level, station assembly would employ astronauts in EVA with help from the RMS which would be controlled directly by EVA astronauts or those in the cabin. Many of the initial modules could be assembled on earth or in the shuttle bay area. The astronauts would have available limited voice input to the system and a CAD/CAM terminal. They would be able to make visual inspections either directly in EVA or indirectly by remote TV camera.

By year 2000, the assembly process could be far less structured with diverse assembly objectives using large and small robots. The OMV would contain 2 arms operating near a staging area containing spare parts, a CAD/CAE terminal, a measurement module for dimensional and alignment checks, smart tooling and clamping pallets, and some tools for small scale machining and forming on site. The OMV would have the capability for micro-positioning as well as measurement over larger distances.

By year 2010, the station assembly process will be similar to that now undertaken in remote operations on earth except that much of the work will be performed by robots. The assembly will be more complex and varied requiring adaptive robots capable of treating detailing and lack of structure. The number of different parts on the premises of the station can never be large because of the penalty of weight and expense. Hence, many of these parts will be altered on site using such unit processes as:

- resistance spot welding
- electron beam welding for continuous joint and seal forming
- localized light forming
- precision light machining; cutting, riveting, drilling, routing, etc.

This means that there will not only be a necessary improvement in the level of sensing and perception but also a dramatic improvement in the mechanical control of the arm structure to maintain precision under process disturbances so that direct commands to the robot can be taken from the on-board database and CAD/CAE system. Otherwise, each of the different unit processes listed above would require a unique dedicated machine—a very costly and heavy alternative to a fully integrated self-contained generic robot capable of all those functions simply by interchanging specialized lightweight end-effectors.

Martin Marietta (1984) conceptualized an extension of the 60 ft RMS by Spar Corporation in the form of a versatile cherry picker (see Figs. 3 and 4). This module would weigh about 600 lbs., would be 4 ft. in diameter when stowed, and have a 10-year life with maintenance. It would have two 7-DOF 50-inch reach arms capable of carrying a load of 50 lbs at 18 in/sec. It also would have a 3 lb tip force for backdriveability, a 1.0% total deflection under load (0.5 inch), and a level of backlash of 0.2% (0.1 inch). Obviously,

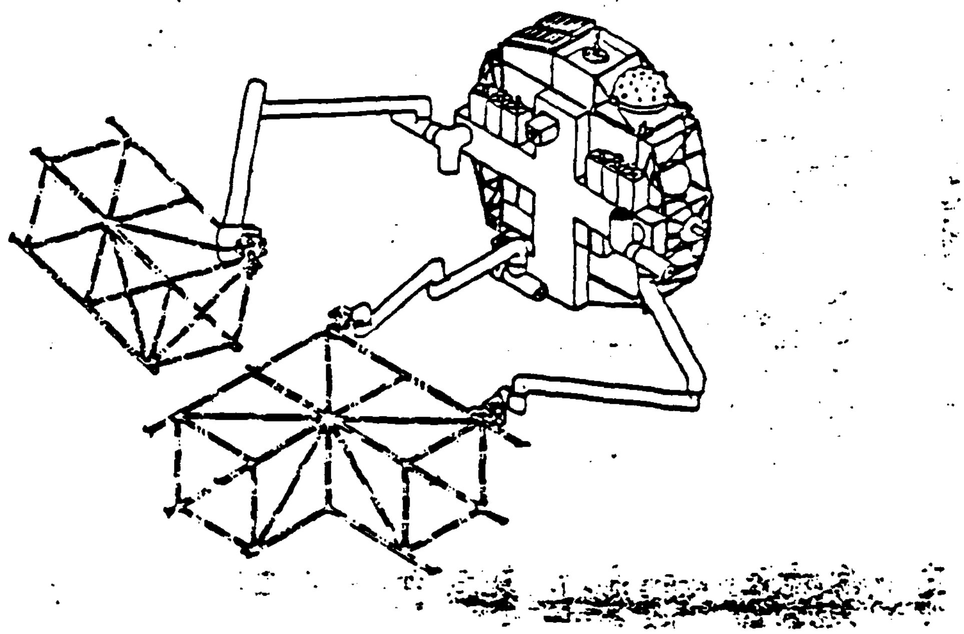


FIG 3. CONCEPT FOR SPACE STATION ASSEMBLY (NASA MEMO 87566)

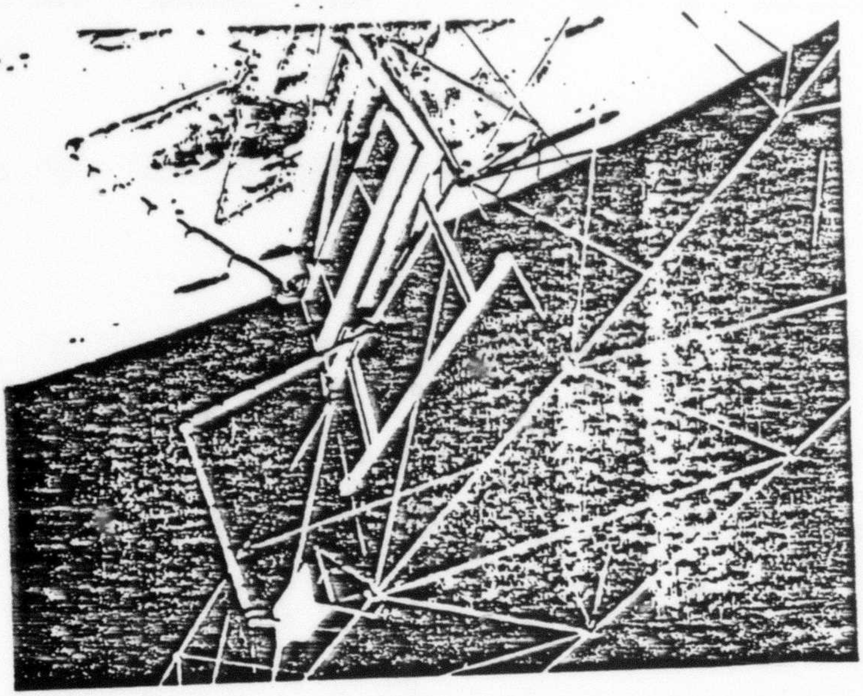


FIG. 4 CHERRY PICKER CONCEPT FOR SPACE STATION ASSEMBLY (MARTIN MARIETTA 1984)

this technology is suitable for only the simplest low precision tasks appropriate to the IOC.

Beyond this point, cooperating robots and walking systems will become essential for general assembly in conjunction with the precision light machining robot. As mentioned in Sec. II, these next generation systems will ask a great deal of the research community. Hence, a major R&D program should be established now to make demonstration of this class of technology likely by 1995 with implementation in the space station by 2000.

IV. SPACE STATION MAINTENANCE AND REPAIR

The space station is without doubt a highly complex one-of-a-kind concept which will require continuous inspection for fatigue failures, flaws, meteorite damage, etc. The structure will necessarily be made up of very lightweight and specialized materials operating in high radiation, thermal shocks, and unwanted but likely cyclic vibrations. The analogy with the maintenance of nuclear reactors is useful as a guide to what can be expected for the space station. Originally, nuclear reactors were intended to be built with such quality that they would rarely fail, making repairs necessary only infrequently. Unfortunately, they have the same outage rate as coal and petroleum fired plants with at least 40% of the downtime due to forced (unplanned) outages. The steam generator for PWR's and the valve for BWR's are major sources of this downtime.

In the case of the space shuttle, 4 to 6 failures occur per day of operation which is considered a major drain on crew time--hence, the need for robots to reduce this time pressure. Some of the component failures that do occur are in:

- avionics
- thermal shield
- brakes
- landing gear
- TV cameras
- recorders
- printers
- door actuators
- fluid coolant leaks
- heaters
- thermostats
- circuit breakers
- switches
- robot actuator control
- aging composite materials

The IOC space station repair will use telerobots to inspect the hull and structure with follow-up by EVA astronauts. A laser beam in the end-effector will look for cracks, holes, surface blemishes, etc. End-effector modules could use material dependent contact probes assisted by high resolution machine vision. These inspections would occur at low velocity over the solar array, thermal radiator, hulls, windows, mirrors, etc., using the CAD database to avoid major obstacles. Once a flaw has been located, it would be

reinspected by a high resolution system to categorize it and quantify it sufficiently to make development of a repair plan possible. This plan would involve the use of special end-effector tools for sealants, adhesives, patches, etc., performed either by EVA astronauts or by mobile robots.

By 2000, the flaw inspection routine would be highly automated and continuous. It would be generalized to include more complex structures defined in the CAD/CAM database. Most common structural flaws would be automatically repaired by using specialized end-effectors that incorporate a fine positioning system.

Finally, by year 2010, flaws would no longer be patched except as a temporary measure. This more inclusive procedure would require the removal of large scale panels and substructures for repair by autonomous and adaptive mobile robot systems. This level of implementation would allow the removal of the thermal shield, localized machining to remove a flaw, welding to reestablish the structure, and replacement of the thermal shield by spray-on foam.

Because of the unpredictable nature of maintenance and repair, the level of technology in the robot to deal with this uncertainty must be higher than that required for the assembly of the station itself. This means that the robot must be very versatile and adaptive, capable of a very wide range of functions on demand and in emergencies. This problem is somewhat similar to the DoD logistics field maintenance and repair situation. It is recommended that NASA work closely with the DoD logistics community to unify the development of the required technology.

V. SATELLITE SERVICING AND REPAIR

One of the major functions of the space station will be to act as a service station and depot to supply fuel and to perform maintenance of satellites deployed either for civil sector or defense purposes. For example, it is claimed that the $125 million seasat (satellite) could have been saved using the resources of a space station. It is projected that 80% of the space station activity will be associated with satellite servicing including:

maintenance & repair
refurbishment
resupply
refueling
cleaning and storage of space debris

GE (1984) suggested that 75 missions would occur per year involving 2500 crew hours with 2/3 IVA and 1/3 EVA. They note that the existing zigzag earth-to-station communications will require 5 to 10 seconds. This time delay is far too long for earth-to-space station robot control. On the other hand, direct line of sight (4000 km) would give a time lapse of less than 30 msec, but the operational window would be one to four hours making time of the essence. This 30 msec time lapse is the maximum that could be allowed for real time dynamic control (including, of course, computational time as well).

Another important issue for satellite servicing is the long docking time

of 8 to 10 hours which may be reduced to 4 to 7 hours with improved docking mechanism technology. Satellites develop over time a "wobble" on top of their desired stabilizing spin. This motion must be assessed by sensors which provide information so the computer can develop a capture plan. Then the conceptualized service module must "rigidize" its attachment to the satellite so that it can perform its precise maintenance operations. This step must be undertaken with great care to prevent damage. Because of the complex relative motions of the two systems (the satellite and service module) use of generalized robot arms (2 or more) to capture the satellite would require a very high level of dynamic modeling and control (a level of technology not available in the near term).

It will be most economical to repair the satellite by the service module in-situ. Otherwise, it will have to be taken to the space station (see Fig. 5) for major repairs (a task of the same level of complexity as repairing the space station itself—see Sec. III).

The IOC for satellite servicing will involve a service module (including a manipulator) capable of determining the actual motion of the satellite and planning its capture (see Fig. 6). A thruster pack will then be attached by the manipulator to the satellite at a built-in hard point which will then be used to stabilize it (a very demanding operation). Then these packs will be returned to the OMV using the manipulator. A remote mobile manipulator is then attached to the satellite along with fixtures with special features to enhance repair. Damaged modules (from meteorites or from failure) are removed by a series of small unit operations (screwing, bolting, clamping, winding, coiling, locking, etc.). Replacement modules may have to come from the space station or from earth. These operations imply the need for quick-change tools at the end-effector, man-machine interface for robot control, and a high level of mobility in the service module.

By year 2000, the service module would have its own on-board intelligence using natural language commands, automated planning, detailed machine vision, sophisticated dynamic control of manipulators, and multiple arms to handle delicate assembly tasks. A limited capability to perform in-situ precision machining and welding for damage of the satellite would be possible with the service robot system.

By year 2010, the satellite service module would be capable of performing most repair operations in-situ. There would be on-board an autonomous CAD/CAE database operating two or more inspection and service robots capable of a broad range of precision light machining tasks. The service module would also have some capacity to repair itself meaning that one robot could undertake the repair of another on-board robot manipulator.

The primary constraint on the service module will be weight and its limited collection of spare parts, modules, or end-effector tools. Because of this constraint, the robot manipulators, sensors, and on-board intelligence will have to be as self-contained as possible with versatile unit processes of light machining, welding, forming, soldering, etc., at various scales (the scale of the computer chip to the scale of an actuator prime mover).

Generally, this technology is a subset of the technology on-board the space station itself but because it must be lightweight and responsive to unstructured tasks, the robot technology must be relatively more advanced.

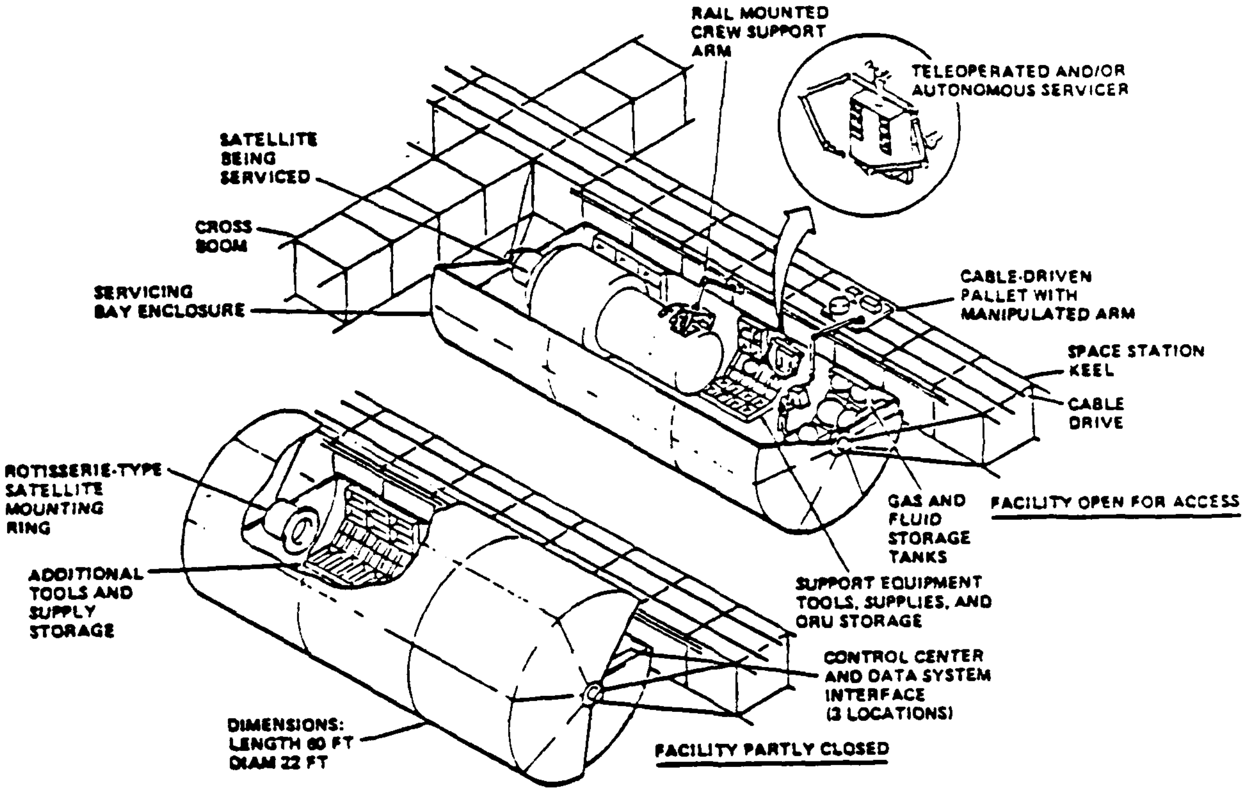


FIG. 5 ENCLOSED SERVICE BAY CONCEPT FOR SPACE STATION (TRW, 1984)

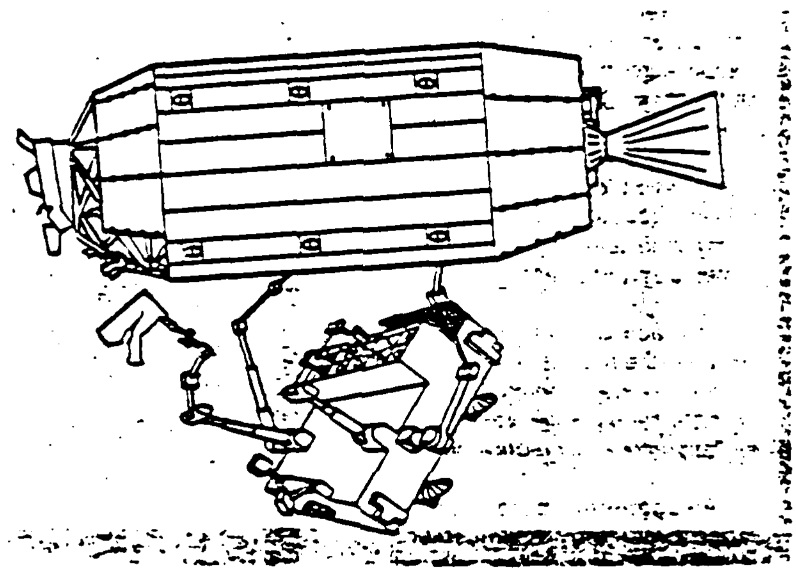


FIG. 6 REMOTE SATELLITE SERVICING VEHICLE CONCEPT (NASA MEMO 87566)

This will put unique demands on the research community to develop the needed technology (see Sec. VIII).

VI. HAZARDOUS MANUFACTURING AND LABORATORY EXPERIMENTS

The economic foundation to perform manufacturing in space is not assured and, by comparison to satellite servicing, is not considered a top priority of the station. Nonetheless, the unique micro-g environment, if extremely stable, invites serious consideration of long duration (hours to days) manufacturing processes and unique laboratory experiments. Some of these processes and experiments may be hazardous to humans or the processes would be unable to tolerate the slightest amount of contamination (the ultimate clean room) from astronauts. Hence, these facilities are perceived to be modules which will "never" be invaded directly by the astronaut partly because of the disturbance to the module's stability that might occur. In fact, it is recommended that these special lab modules be suspended from the space station platform in much the same way that seismographs are suspended. This isolates the module from the shock and resulting oscillations of docking at the station (more than 75 times/year). When access to the module by robots and supplies is required, it can be temporarily rigidized.

Lab experiments in space must necessarily be open-ended and unpredictable to be of value to the researcher which means that an astronaut as a non-specialist can contribute little to the experiment itself. This means that the experimental equipment will generally be rather delicate and need frequent monitoring and adjustment (partially because of the weight penalty).

Furnaces, power sources, control units, data-acquisition instruments, inventory and warehousing systems, etc. are all dedicated subsystems which weigh a great deal, increase the number of potential unique failures, and require specialized maintenance functions. Hence, whenever possible, it is recommended that use be made of "generic" multi-purpose machines such as robots where a large variety of specialized end-effectors make a spectrum of functions feasible.

An analogous system on earth is the secure automated fuel (engineering) plant (SAFE)* being designed to use robots to manufacture nuclear fuel pellets with a high level of security by rarely requiring access by humans. This plant carefully stages the manufacturing process and monitors each stage. Its highly structured form allows simple but extremely reliable robots to perform all the flexible (changeable) tasks. Hence, the only need for dexterous robots is either for maintenance or emergencies. A similar structure is recommended for the manufacturing and experimental lab modules.

Another potential hazard is the possible need for a nuclear power source for the station. This involves radiation, high temperature and pressure, and some chemicals. Such systems must be maintained since failure is probable because of their complexity. Because of the penalty of weight, they will not be as well shielded as on earth. In this case, the service robot(s) would be part of the power station because once the robot is contaminated it could not be brought back to a maintenance bay for repair.

---

* See Ref. 11.

In the IOC stage of the space station development, simple processing of earth provided materials in micro-g gravity would be pursued. Electron beam welding and ion-beam implantation in a contaminant-free environment would be investigated. The lab experiments may include free fall to aid in separation in bio-organic systems and the establishment of very uniform thin films. No human maintenance of these specialized systems would be allowed which would require a fairly dexterous, lightweight, mobile, multi-purpose manipulator system perhaps partially under human control by telepresence from the cabin. These maintenance robots cannot cause any vibrations in the lab module (a level of smoothness not usually seen in robots especially since shock to the module can be transmitted through the base of the robot as well).

Again, because of the penalty of weight, the lab module is likely to be very compact, which means that its volume will be relatively full of obstacles to be traversed by the robot. Generally, an obstacle strewn environment means that the robot must be redundant (have several extra DOF) to allow it to not only get access to a task but also to perform complex spatial motions while in that constrained space. It strongly suggests a modular snake configuration for the robot. This means that modules (extra DOF) could be added or discarded on the basis of need to enhance the level of dexterity and obstacle avoidance needed. Unfortunately, the required design capability for a modular snake robot probably would have to be built by NASA itself. Furthermore, the operational software based on complex issues of geometry, dynamic modeling, control, and AI would also have to be established by NASA.

GE (1984) recommended that the manufacturing facility (see Fig. 7) be able to perform the following specialized tasks

slicing
polishing
cleaning
sawing
separation
ion implant
photo resist
annealing
E-beam direct write
reactive ion etch

One of the opportunities of such a self-contained module is the manufacture of computer chips using the hazardous material galium arsenide. It is said that such ultra pure computer chips could lead to the supercomputer on a chip.

By year 2000, the manufacturing and experimental lab facilities would employ a force reflecting and "electronically rigid" robot for maintenance of the modules. This robot may, from time-to-time, be required to perform precision welding, precision light machining, and other disturbance related tasks in order to minimize the number of maintenance systems to treat the increasing complexity in the modules themselves.

By the year 2010, these facilities could be operated almost completely by off-line programming of electronically rigid handling and maintenance robots. The most demanding operation in the manufacturing facility could now be treated—the cleaning and renovation of the solid-state materials furnace after solidification following a power failure. Without this capability, the

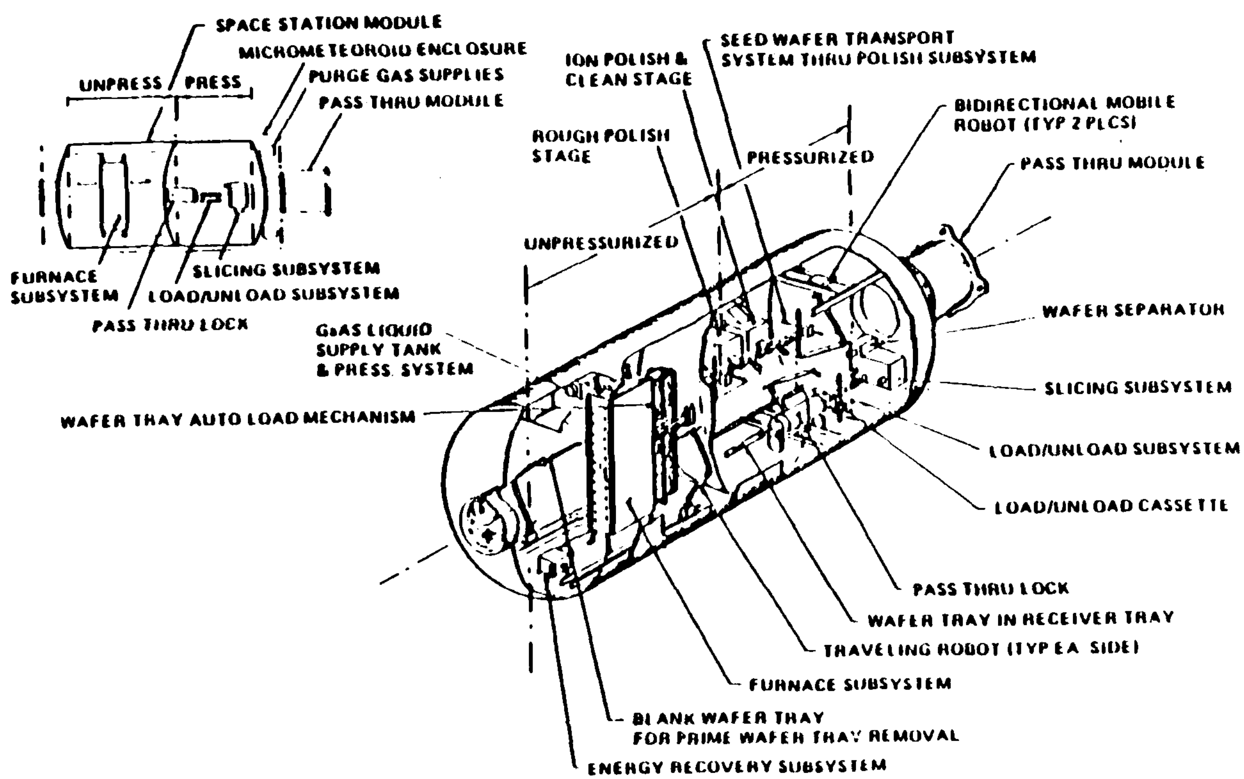


FIG. 7 CRYSTAL AND MANUFACTURING CONCEPT (GE, 1984)

value of the whole module could be destroyed. Because of this critical situation, a standby furnace should probably be provided until remote renovation is assured.

VII. MAINTENANCE OF ROBOTS

Robots in space will generally be modular, lightweight, intricate, and in some cases, delicate. The first purpose of the robot is to perform its intended function. Its maintainability will be secondary, but modularity can perhaps make the best of both requirements. They will be unavailable (2% downtime would be quite low because of their generic complexity, lightness, etc.) about 2/3 of the time for regular maintenance and 1/3 of the time for forced, emergency maintenance. They will have to be up-dated with new modules for improved technology. Infrequently, overloads from docking shock could deform the robot structure itself. Overall, the maintenance of maintenance robots will prove to be as important to the space station as other critical operations.

In order to deal with the increase in complexity of the space station, the level of complexity (see Fig. 8) of the supporting robot systems will also increase. Generally, the robot system will tend to degenerate with use (thousands of hours) and its system parameters will change. Some of these changes can be dealt with directly by self-diagnosis and corrections to the operating software (a very high level of machine intelligence). But a robot is hardly likely to repair its structural hardware by itself. Consequently, a certain duality may become necessary to keep the whole system operational.

VIII. DESCRIPTION OF COMPONENT TECHNOLOGIES FOR ROBOTIC SYSTEMS FOR THE SPACE STATION

Fourteen component technologies have been identified which adequately represent the total robot system. These 14 component technologies have been described in greater detail in a companion assessment* for applications appropriate to the fields of manufacturing, energy, military, and human augmentation and agriculture. Here, they will be described in terms of the special needs associated with the operation and maintenance of the space station.

1. Structural Geometry

Two dominant issues will effect the geometric design of robotic structures for use in the space station. They must be very light weight yet they must operate in the unique and unforgiving condition of the micro-g gravity field. Many applications will require that significant end-effector forces be developed while maintaining a high level of precision (i.e., light

---

* See Ref. 10.

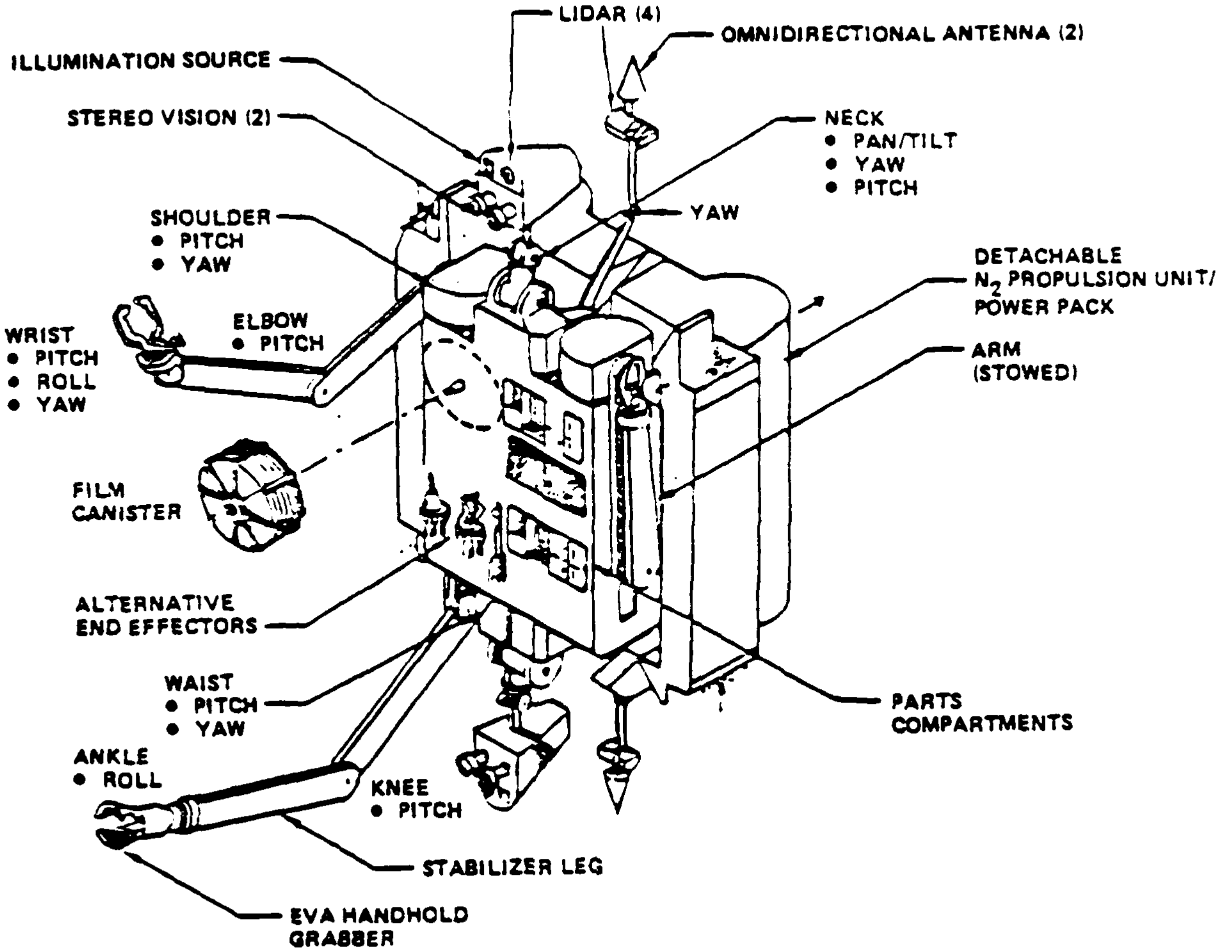


FIG. 8 COMPLEXITY OF SERVICE MODULE SHOWING NEED TO BE ABLE TO PROVIDE MAINTENANCE BY ROBOTS (BOEING, 1984)

machining). On the other hand the crowded environment of compact lab modules will require a higher level of dexterity than is now found in industrial robots. This leads to the question of serial geometry (for high dexterity to perform complex tasks among obstacles) versus parallel geometry (for precision and low weight). Part of the solution may come from modular robot structures which may be configured on demand to best meet the dexterity requirements of a given range of tasks. When precision is a top priority, disturbance rejection can be achieved by making the structure "electronically rigid." This can be best achieved by using a new mechanical architecture of layered large and small prime movers combined with similarly scaled control technologies (which may be called control-in-the-small). In order to achieve this level of technology, the complete dynamic model of the system must be calculated in real time, a requirement not yet met for any existing industrial robot.

By comparison, a much more difficult problem for the motion control of a general serial robot structure is the calculation of the joint values at the actuators to correspond to desired end-effector location. For the general 6 DOF* system, this results in a 32nd order polynomial of great complexity. Going to 7 DOF or more results in a motion problem which has no deterministic solution. This means that the extra DOF (redundant) system must be treated by some form of artificial intelligence technique for internal decision making. Problems of this class are only being conceptualized by research teams today.

The master controller for the space station man-machine interface will have to be geometrically optimized for best match to the human operator and universal in that one geometry is capable of driving any "slave" robot. Docking mechanisms will require special geometries which are highly parallel, dynamically responsive, and easily rigidized even though light weight. Because of the micro-g gravity field, many operations will require two arms operating on the same object; i.e., twelve actuators working to control 6 to 8 output motions (a redundancy of 4 to 6 input parameters). This problem is of great complexity and is presently only being formulated in research labs. A similar problem involves a snake robot which is composed of a series of 2 DOF and 3 DOF modules to create systems of 8 or more DOF of very high dexterity. A problem of even greater complexity involves the motion of walking machines (a proposed mobility platform for the space station). In the case of 4 actuators in each of 6 legs, this would lead to 24 inputs trying to control 6 outputs (a redundancy of 18). The solution will entail a combination of modern control theory, artificial intelligence, structural dynamics, and analytical geometry all operating in real time (less than 30 msec.).

2. Structural Dynamics

Because of the special constraint of minimum weight, an extraordinary effort to develop a full fly-by-wire approach to the dynamic operation of robotic systems for the space station will be essential. In some cases, massive inertia loads will be involved. Or the existing 55 ft. long Robot Manipulator System (RMS) in the shuttle could be used with a smaller robot to form a "cherry picker" configuration. Because of the large scale of these

*Degree-of-Freedom

systems, dynamic collision avoidance procedures must be imbedded in the operating software. Also, robot structures may need to operate in an acceleration field during docking maneuvers. This field creates a very complex inertia load distribution which has to be dealt with by real time dynamic modeling and adaptive control techniques to ensure that the required docking forces and berthing dynamics are being maintained. On a smaller physical scale, many unit processes during assembly and maintenance will involve light machining tasks which generate large force disturbances.

The overall goal to create a robot structure of minimum weight results in links which are very flexible and actuators which have marginal load capacities. These realities put an exceptional demand on the operational control software which must be used to compensate for these limitations. The operational software can be established only by maintaining an inclusive dynamic model of the robot structure (i.e., the fly-by-wire concept now used in the control of modern fighter aircraft). This software must function in real time which, in this case, means at less than 30 msec. (This has been done for 15 DOF systems of completely general geometry and mass content.). Given the dynamic model, it is possible to maintain a given end-effector force, a desired level of dynamic response, or to compensate for deformations due to force disturbances. The ability to achieve these goals is complicated when extra DOF are employed to provide additional dexterity (i.e., in the form of a snake robot). A combination of graphics (CAD) and control based on the dynamic model and artificial intelligence (AI) for internal decision making will be required to operate these redundant systems.

Probably the most complex system to dynamically control is the walking (or crawling) system. This entails the balancing of internal dynamic forces (at each of up to 24 input actuators) to create 6 desired output forces acting on the body of the walking machine. Should the body also carry a robot manipulator to act on its surroundings, a much higher level of dynamic modeling and control would then be required.

3. Prime Movers

Again, the principal issue affecting the selection of actuators is that they must be of minimum weight and maximum load capacity. Precision operations under load disturbances also require a very stiff driver—i.e., it is not easily dislocated from its reference position. In addition, simplicity in the bearings would be highly desirable (ceramic bearing surfaces). The driver should involve a minimal amount of backlash, stiction, and hysteresis loss to best work in the micro-g gravity field. One of the suggestions has been to employ direct drive motors because of their simplicity. Unfortunately, the penalty of their inherent high weight and low stiffness will probably make them unworkable for most space applications. By contrast, the harmonic drive gives a very high load capacity, stiffness, and responsiveness in a very compact and light weight package.

The conclusion must be that prime movers and actuator modules are an important subject for development to meet special space station objectives. Generally, actuators can be designed for lower load capacity if they are integrated into parallel structures. Mixed large and small actuators in a layered structure can mean that the principal actuators can be smaller since the small actuators are dedicated to the precision requirements of the system. Many have suggested piezoelectric drivers, although their motion range and power capacity is severely limited.

4. Actuator Modules

The concept of actuator modules involves the compact arrangement of 1,2, or 3 DOF structures with as many inputs to create subsystems similar to the human elbow, ankle, knuckle, wrist, and shoulder all easily scaled up and down for various uses in robot systems. Especially for use in space, these modules should contain their own dedicated sensors and operational software so that they can be replaced easily during maintenance or during system up-dates. This level of modularity would minimize the number of separate units to be maintained or inventoried in the space station. Parallelism in the geometry of the module reduces the total weight (the example of the Odetics parallel walking machine saves a factor of 10 or more). Antagonism of input actuators as seen in the biological system provides a high level of resolution while maintaining a high load capacity. The structural parallelism and antagonism for the motion control of the human eye is indicative of all the correct characteristics to be found in actuator modules designed for implementation in space.

Such modules can be quickly assembled into a serial robot, a parallel robot, a 10 DOF snake, a miniature robot, a cherry picker configuration, or a docking robot system. Once this level of modularity is reached, each robot can quickly be reconfigured to meet a specific range of tasks having a need for dexterity, or precision, or reach, or.... Other specialized modules such as a micro-manipulator for vernier adjustments in 6 DOF are also of this class. It is recommended that NASA build a generic collection of such modules easily scaled and interfaced before it invests heavily in the robotic systems technology that will be required.

5. End-Effectors

The end-effectors are either the specialized tools attached to the end plate of the robot (drilling, routing, welding, etc.) or the dexterous multi-fingered hands which allow general manipulation of the work in progess. Specialized tools of the type being considered for the space station are either being developed or can be found in other applications such as remote maintenance of nuclear reactors. These tools must be relied on to perform precision operations which must be performed during assembly or as the result of damage to the station. Frequently, the handling of small and delicate objects will be necessary to perform laboratory experiments, to do remote satellite maintenance, or to reassemble a satellite. The variety of these smaller objects will require the use of a conforming dexterous end-effector, usually conceptualized as a multi-fingered hand. Such a hand will not allow for precision positioning without the human's hand-eye coordination as part of the control loop. Hence, the use of dexterous hands will be demanding of the crew's time and should be considered primarily as a last resort and for special events.

One of the considered needs for assembly is a variable compliance capability in the end-effector to assist certain forms of slip-fit joining of parts. Generally, high compliance can be used when precision is already built into the parts. The opposite (high stiffness) generally means that precision operations (i.e., light machining) can be undertaken. Hence, the governing software must be able to adjust the system to either be passively compliant or electronically rigid.

The fingers of this dexterous hand may profitably employ fiber-optic eyes for inspection purposes and tactile surfaces to help identify objects when either direct or indirect vision is unavailable. Tactile arrays have yet to be employed to do more than provide threshold information about an object communicated directly to a computer or to a visual display to the operator.

6. Graphics/CAD

A reality of the space station which deserves early attention is the need for near optimal design of a facility for which there will only be one prototype. Hence, computer-aided-design (CAD) must be used in every aspect of the system's design and its operation. This is particularly true in robotics, where most industrial systems are designed before the tools of CAD are employed to their full effect. The result of the CAD effort will be a complete data base about the "as designed" space station which must be up-dated to account for the "as built" system. Once such a data base exists, it can be used to plan for various tasks either by simulation or by semi-automatic planning using principles of artificial intelligence. The as built data base becomes critical when either unproven maintenance activity is being undertaken or when responding to emergencies. For example, it may be possible to superimpose actual visual feedback with the stored data base scene where differences may be isolated either computationally or by the human operator. The data base may be used to display deformations in the robot structure by color coding its surface to visually inform the operator about the condition of the robot.

One of the major issues for the maintenance of satellites in-situ is the long time delays for a cycle of communication (up to 0.5 sec.). It is proposed to use a predictive display (a ghosted robot) from the data base to smooth out the visual feedback to the operator. This type of technology would be essential anyway to train future astronauts (as does the Link trainer for aircraft pilots) and to develop helmet displays which could be used in EVA.

7. Sensor Technology

The most important physical parameter to be sensed in robotic operations is the force level being experienced throughout the robot structure (its end-effector, joints, links, and base). This level of information must be known accurately in order to make corrections to the command signals so that precision tasks can be performed even with large force variations. The end-effector force can now be measured by standard light weight 6 DOF force sensor modules. Accurate force information can then be transmitted to the operational software or to the human operator. Also, current flow at electrical actuators can be used to provide local load information throughout the structure. Since the primary force generator in space will be inertial, velocity and acceleration transducers might prove helpful, especially during docking procedures.

Remote operations will involve questions of collision avoidance. In some cases, the data base may prove inadequate or in error. Then, collision sensors may be helpful in preventing major damage. Unfortunately, such sensors can not be expected to provide more than the simplest of "threshold" signals. Tactile sensing is somewhat of the same character (see previous section).

Because of the high level of deformation expected in space station robotic systems (due to low weight requirements), a much higher need occurs to obtain accurate positional data on the location of the end-effector in world coordinates. Laser systems offer promise but no proven technology exists at this time. Also, specialized digital software for machine vision to monitor a finite number of end-effector "spots" is now being developed.

As with prime movers, sensors should be designed as self-contained modules with much of their own software for easy interfacing with the rest of the system. This modularity is basic to rapid replacement for maintenance or up-grading.

8. Vision

The first use of vision for the space station will be analog feedback to the human operator of the robotic structure. In this regard, visual enhancement by mathematical filtering would be very useful especially when comparing the actual scene with that stored in the data base. The principal use of vision, however, will be to perform automatic inspection of the space station itself. This inspection will be done continuously by looking for flaws or damage without human intervention until some anomaly relative to the data base is located at which time human inspection would be called for. In addition, vision will be useful to minimize the danger of collision with space station components or with other robots. As mentioned earlier, a predictive display based on a programmed motion for the robot could be combined with vision to smooth out operator feedback when distance related time delays occur.

One of the most demanding tasks involved in the space station is docking remotely powered vehicles with satellites in order to perform maintenance or resupply. This docking procedure now requires 8 to 10 hours. It is proposed to use vision technology to mathematically describe the actual motion of the satellite relative to the RPV, use AI principles to plan its capture, and dynamically program the RPV manipulators to physically grasp the satellite with minimal reaction forces.

Finally, vision will be a major component in any training facility that would be essential to prepare astronauts for the spectrum of complex tasks to be performed in space station activity—both routine and forced (due to emergencies). This training facility will greatly resemble the Link trainer system now being used to train aircraft pilots.

9. Artificial Intelligence

Artificial intelligence is a valuable tool to treat non-deterministic decision making or resource distribution as an aid to the human operator. The goal is to reduce the burden on the operator by autonomously carrying out routine inspection and maintenance tasks where infrequent supervisory intervention by the operator would be necessary. This means that trajectory planning, collision avoidance, fault identification, and task prioritization would be the responsibility of the machine intelligence on board the robot system. This level of decision making may be applied to the robot itself in order to monitor functional degradation, locate defective modules, plan for its own maintenance, and instruct a neighboring robot to carry out the plan.

Internally, robot manipulators that are highly deterministic usually have 6 inputs to develop 6 outputs. Should more inputs (say 10) exist, the system becomes redundant (4 extra DOF) and uncontrollable by standard techniques now used for industrial robots. The array of 10 inputs must be balanced (in terms of force, speed, energy, power, etc.) in order to carry out the desired task at the end-effector. Hence, AI principles will be necessary to evaluate the task, determine if the robot should be reconfigured (dimensions changed, modules added or removed, if larger load capacity is required, etc.), assess the level of precision required, employ disturbance rejection software if needed, etc. Hence, a very high level of AI must be developed to properly employ a generic, modular, precise robotic structure in space station operations. The type of AI that is actually required, however, will be dependent on technology base issues found in the electrical and mechanical engineering fields.

10. Intelligent Control

Intelligent control encompasses all deterministic techniques used in the operational software of the robot which is used to enhance its precision, speed, smoothness, disturbance rejection by modern control methods, digital control, adaptive control, etc. The control software must be defined in terms of a complete parametric description of the system's link dimensions, deformation rates, mass content, and actuator control parameters. Such a detailed parametric identification can only be achieved by a high level of metrology—a technique now being formulated in research labs. In fact, since some of these parameters will change over the life of the machine, some aspects of this metrology may need to be on board the space station itself, especially when modules are either replaced or interchanged.

Intelligent control then depends on a complete and accurate analytical model which must be calculated in real time (less than 30 msec.). This model can then be used to train astronauts, to perfect dynamic collision avoidance techniques, docking procedures, compliance control, etc. Layered control of two or more scales can create a hierarchical architecture for the system hardware and software known as control-in-the-small which is much more effective in providing "feedforward compensation" to reject force disturbances from such tasks as light machining. This sophistication is warranted because of the high level of deformations which result since the robot must be as light as possible. Elimination of these large deflections not only makes precision tasks feasible, it also significantly reduces confusion, fatigue, and frustration of the operator when he is trying to perform this type of task by manual control.

Beyond this level of understanding, there will be a concern for the level of reaction forces (shock level) transmitted through the base of the robot to the space station structure. These forces could easily disturb the environment of critical laboratory operations. Also, criteria must be established for the balance of 12 actuator forces (and other parameters) to create 6 desired force components on objects held jointly by two robots (dual robots). This dual robot problem is quickly expanded to one of 6 parallel acting robots when developing the operational software of a 6 legged walking machine (24 or more inputs to create 6 outputs). Each of the forces at the feet of these legs must be controlled in order to not disturb or deform the space station structure while it walks. This problem is the most advanced form of intelligent control and will require a major theoretical development.

11. Software Modules

Because of the need to use a minimum number of distinct components to build up the consort of robot systems to be used in the space station and its operation, hardware modularity will greatly reduce the size of the on-board inventory (a high priority) to maintain or up-date this technology. This means that the software will also have to be highly modularized to match this special architecture. This then allows for the addition of joints, the changing of link dimensions, the increase or decrease in compliance, etc. The following is a partial listing of such modules:

actuator control
sensors data reduction
end-effector operation
special tools (drilling, welding, etc.)
docking procedures
satellite motion identification software
fault isolation technology
vision control
local servo motor control
force sensors
micro-manipulator control
tactile array sensor software

Also, it is recommended that much of the system software be modularized so that it can easily be maintained or up-dated.

12. Computer Architecture

The pace of development of computer technology makes the whole concept of the space station feasible. The pressing reality for industrial robotics is that thus far, computer integration has occurred only to satisfy the most undemanding unit processes (painting, handling, welding, etc.) or applications (low value-added operations in manufacturing). Consequently, a broad based effort by NASA to more completely integrate computer and control technology with generic mechanical architecture is not only essential for space station operations but also of real potential value in significantly enhancing productivity in manufacturing.

The promise of computer technology for space station automation and robotics is based on the broad spectrum of these technologies from the component level to the system level:

VHSIC chips
Arithmetics
Array processors
Mini-computers
Super-computers

This collection suggests that no computational needs in robotics should go unattended since all components in the robot system (sensors, actuators, structure...) can now be brought to a much higher functional level. There should be no reluctance to match the architectures of all components to the wide availability of computer architecture and vice versa.

Much of this need to integrate computer technologies has been described elsewhere in this paper. Some of these key areas will be listed here for completeness. Because of the requirements for low weight, versatility to respond to unknown tasks, autonomy to carry out continuous inspection, precision in unit processes such as drilling, routing, forming, etc., collision avoidance in a complex and changing environment, access by humans for intervention and supervisory control, etc.; the level of computer integration will have to far exceed any previous effort and will require extraordinary care in structuring the research and development program.

In every case, direct support must be maintained in the form of accurate numerical documentation from the data base--a level of information never attempted before for robotics. This means that all activity should be quantified and programmed to minimize the level of uncertainty. Uncertainty in the operation of the robot should be accepted only when the benefits are very high, i.e., for collision avoidance, high levels of dexterity to carry out complex operations, etc. This level of uncertainty and the associated need to employ principles of artificial intelligence becomes pervasive in dual arm operations, walking machines, docking operations, automated inspection, motion planning in cluttered environments, etc.

On the other end of the development program, the need to design the complex hardware and software for this advanced robotics technology must be dealt with. Thus far, only minimal efforts to develop a technology base for robot system design has been pursued. Essentially, the CAD technology must preceed the operational technology. Having the CAD graphics capacity, it can be used to develop the required training facilities for astronauts to prepare for space station activity.

13. Communication Interfaces

The space station may be thought of as having the same array of interfaces as would be found in a modern factory. The highly desirable feature recommended for space station systems is modularity to enhance maintenance and technological up-dates. The more modular the space station and its supporting systems (robotics), the more concern there must be for interface issues. The most dominant interface is between man and machine but others exist:

Lab subsystems
Astronaut support
Satellite control and maintenance
Sensing and inspection
CAD data base
Hierarchical decision layering
etc.

Unfortunately, some communication delays will occur between the space station and earth, or with satellites, or with RPV's working on satellites, etc. Finite time windows will be available (measured in hours) such that time may be of the essence.

It is recommended that a clear effort be made by NASA very early in the program to establish standards for interfaces at all levels from specialized tools to space station communications between its prinicpal sub-modules. The NBS factory floor interface program may be a real asset in this effort.

14. Man-Machine Interface

The history of complex technologies such as nuclear reactors shows that they are under repair approximately 25% of the time with 40% of this downtime due to forced outages. This lack of availability will potentially occur for the space station principally because of its complexity and because it must be regarded as a prototype system. This reality suggests that the operation and maintenance of the space station will involve the deep participation of the human ability to make decisions based on uncertain information. In fact, even though there is a pressing need for autonomous operation to reduce the burden on the on-board personnel, as the technology becomes more adaptive and more capable of performing complex operations, the ability of the operator to intervene becomes more important—not less. As the technology advances, human decision making (judgement) can enter in at a much higher level.

The need for human intervention is best provided in terms of kinesthetic interface because of the high rates of analog information transfer that is feasible. This interface has been conceptualized as a bilateral force-feedback manual controller. Because astronauts will have to control a large range of unique robots (many which will be reconfigured to meet a given task), the controller must be universal with software capable of driving any robot. This universal nature also reduces the training effort faced by the astronaut.

The "universal" requirement means that the coupling software must operate in real time and be highly adaptable. It must enhance signals to the operator, filter out jitters or gross errors from the operator, perfect global commands such as constraining the end-effector to track the surface of a sphere, to mathematically change orientation, to change scales, to monitor manipulator operations for accidents, impending collisions, overloads, etc. When it becomes necessary for one astronaut to simultaneously control both arms in a dual arm system, then the quality of the interface will become critical. It is obvious that an advanced manual controller would be invaluable as a training aid.

The "universal" aspect of the manual controller also has a significant impact on the design of the man-machine interface. It means that the master (or manual controller) can be optimized for its primary interface with the human operator, it can be made lightweight, and it can be kinesthetically transparent. On the other hand, the slave (the robot manipulator) can then be optimized for its principal range of functions without being compromised by constraints or limitations which would occur from a geometrically similar master-slave combination. Dissimilar geometries means that the software will have to be far more general but doing so provides a total system which is much more adaptable to changing applications.

IX. DESCRIPTION OF CRITERIA FOR THE ADVANCED DEVELOPMENT IN ROBOTICS SYSTEMS TECHNOLOGY FOR THE SPACE STATION

NASA faces an unusually broad development task to move the technology for robotics forward sufficiently in order to maximize the availability of the space station for both civil and military uses. A program to achieve that development is being defined and documented at this time. One of the most

important parts of this plan must be "measures of success" based on sound system criteria which can be used to monitor progress in the technology. Since robotics must be considered an immature technology, the exact criteria to evaluate progress remains uncertain. It is believed that the following 14 criteria is sufficiently broad to enable a management team adequate means to evaluate the contribution and rate of development of various component and system technologies.

1. Multi-task Capability

The number of distinct tasks a given robot system can perform is a dominant consideration in order to reduce the number of such systems necessary to operate the space station. A reduced number of robot systems implies reduced weight and a reduced inventory of replacement parts. Some of the unit processes that must be performed are:

Operate simple mechanisms
    latches, cranks, slides, handles...
Joining and fastening
    fitting, force fit connectors, spot welding, forming, bolting, screwing, locking, coiling, riveting, electron beam welding
Precision machining
    grinding, sanding, brushing, drilling, routing, trimming, cutting
Handling
    parts transfer, limp materials, slippery materials, warehouseing
Automated inspection
    seam tracking, surface flaws, meteorite damage, etc. on solar arrays, thermal radiators, windows, mirrors...

In addition, there are several complex dynamic motion tasks which are either necessary or may be tested for space station operation. These are:

Docking and grappling maneuvers
Reactionless operations
Stabilization by appendage motion
Rigidization
Catching and storing space debri
Throwing and jumping
Dual robot operations

Overall robot systems are intended to carry out some in-depth functions over the long term such as:

Clean room operations in both manufacturing and experimental lab modules
Self measurement of space station dimensions over large distances
Space station assembly
Repair and maintenance throughout the space station itself
Repair and maintenance of satellites in orbit

This breadth of activity shows that robot systems in the space station face a formidable spectrum of physical tasks. This breadth strongly argues against numerous specialized dedicated machines but in favor of generic multi-purpose robots with an ever increasing level of flexibility. The reality of this need

for generic technology may be instrumental in leading the move away from numerous highly dedicated machines now used throughout U.S. manufacturing.

2. Level of Machine Intelligence

The full array of housekeeping chores, inspection, maintenance, and response to emergencies will overload the limited number of personnel on board the space station. Hence, every effort must be made to automate as many of these operations as possible. This objective can only be met by employing a high level of machine intelligence frequently based on principles of artificial intelligence. Because a continuously updated data base will be used for the space station, the work environment will be reasonably well structured. This means that semi-automatic inspection, near optimal trajectory planning, situation assessments, collision avoidance, etc. are feasible for the space station.

The commands from the human operator through the manual controller can be perfected in terms of functional requirements stored in the data base, jitters and gross errors can be mathematically filtered, motions once taught can be repeated without operator involvement, or a ghosted robot can be used to guide the operator in planning motions in an obstacle environment. Since humans can not either make unaided precision measurements or perform precision operations especially under disturbance, machine intelligence can be used to augment the operator's skill. This becomes especially necessary in the operation of dual arm systems and the automatic foothold selection and walking operation of multi-legged structures. Machine intelligence will be used to control extra degrees of freedom, to reconfigure the robot structure to provide more precision or dexterity, to search for system faults and to call for and plan for corrective action.

This level of machine intelligence far exceeds that available today in industrial robots. It will be achieved only with a consistent and long term commitment to a broad based R&D program by NASA.

3. Time Efficient Operation

Because of the high cost for the space station ($8 billion) and because time may be of the essence in emergencies, the time efficient operation of the supporting robotics technology is an important criteria for its design and implementation. The reference would be either the time for the human alone to perform similar functions on earth or the astronaut alone in EVA. The need for productivity is highlighted by the fact that the shuttle has 4 to 6 failures per day and that docking with a satellite now requires 8 to 10 hours. Accurate situation assessments can be carried out numerically without human intervention. The data base and imbedded AI technology can be used to eliminate time consuming trial and error motion trajectory selection in an obstacle strewn environment. Once a complex motion has been selected, it can be taught by human control and repeated automatically by machine intelligence. In this regard the CAD data base can have a high payoff as a basis for astronaut training. In the robot structure itself, high load capacity, precision and resolution, combined with low stiction and backlash can significantly reduce the fatigue and frustration experienced by the human operator.

4. Unstructured Task Level

The numerical documentation for tasks to be performed by robots on the space station will draw on an extensive and continuously updated CAD data base. Many uncertainties will exist because of the differences of "as is" versus "as designed" resulting from imperfect assembly, maintenance, parts replacement and updates, structural damage, etc. Maintenance procedures may not be as prescribed but as necessary. It may be necessary to correct the incomplete or imprecise work performed by dedicated automation equipment. Patching may be necessary in the near term although the goal must be to restore all components to first class condition. Sensors used in automated dimensional inspection will continuously update the CAD data base. The goal is to reduce the level of numerical uncertainty to a minimum. Had this effort been pursued in the case of nuclear reactors, their maintenance by robots would have become much more feasible and certainly less expensive.

5. Geometric Dexterity

The minimum dexterity required to control spatial motion is 6 DOF. Serial structures are more dexterous (in avoiding obstacles, range of motion, ...) than parallel structures. The human arm contains a mixture of 7 serial and parallel DOF and when combined with the shoulder's motion provides 13 DOF, certainly sufficient to perform most dexterous tasks at the human scale. Essentially, compact and cluttered work environments (as in the space station lab modules) require extra DOF (say a total of 8) to make a wide range of motions feasible. It is conceivable to add extra DOF modules to a robot to enhance its dexterity on demand. Unfortunately, no mathematical scheme has been developed to prescribe the motion of a robot having more than 6 DOF. Only principles of AI combined with physical criteria for best overall operation can be expected to treat this problem. To date, this problem has only been conceptualized in research laboratories.

A maximum level of dexterity may be achieved by a slender "snake" robot having 8 to 10 DOF in the large and several fine DOF in the small (the cherry-picker concept). A highly dexterous end-effector may be necessary in grasping space debri, handling small parts in the maintenance of satellites, etc. This high level of geometric complexity far exceeds the dexterity level that can be achieved by any manual controller. Hence, the operator's commands must concentrate on the function to be achieved while the operating software must automatically command the extra degrees of freedom—a level of intelligence which far exceeds that available today.

6. Portability and Mobility

A major issue for the space station is to establish the ability to move about the station to perform planned or emergency repairs or to perform assembly and disassembly tasks involving:

Material transport
Satellite capture
Service orbiting platforms
Inspect station structures
Inspect solar panels
Inspect tension cable telescopes

The questions of on-board power, tethers, on-board intelligence, on-board inventories of replacement parts, etc. must all be dealth with. There are three approaches for mobility for space station operations. These are:

Rail Transport—This requires tracks to be overlaid over the whole platform structure. It carries an increasing weight penalty as the station grows in size. It appears to be difficult because of its limited motion range to avoid all obstacles. Tracks would provide a high level of reliable motion with precision, high load capacity, and very efficient energy usage.

Crawling —This technology would probably involve some form of walking machine either tethered or with its own power package. Its relatively low weight does not grow with the size of the station. It would be fairly energy efficient but would move slowly and involves a very high level of undeveloped machine intelligence to govern the motion of its legs.

Free Flight —This system would carry its own power package to maneuver by thrusters or by jumping from one part of the station to the next. This method is relatively energy intensive, involves time consuming docking and rigidization procedures, is slow, and would have a low load capacity. It is less reliable than the other concepts and would potentially create problems with its thruster plumes. Overall, it is the most near term technology available and carries a fixed weight penalty when ignoring its fuel consumption.

In the full development of the space station, some balanced combination of all three of these concepts will probably be employed. Hence, it is recommended that they all be developed further for space station evaluation and possible implementation.

7. Precision

Many operations in the space station will require high levels of precision (1 to 10 thousands of an inch) even when the robot structure is disturbed by forces generated by the process being performed. In order to appreciate this requirement, the example of the 5500 lb. Cincinnati Milacron T3-776 industrial robot deforms 20 times (0.200 inch) its resolution (0.010 inch) under its payload of 150 lb. Space station robots will necessarily weigh 1/20 of this robot in order to create a force of 75 lb. In order to achieve the level of precision desired, it will be necessary to make the robot "electronically rigid", a development objective now underway in some research laboratories.

Process disturbances occur from such unit processes as cutting, routing, bending, drilling, force fit assembly, etc. The availability of these processes would significantly reduce the otherwise large inventory of parts that would be required to repair major space station damage (however infrequent). It further simplifies the overall design of the station probably also decreasing its weight.

Precision light machining by generic robots would reduce the number of heavy dedicated machines required to perform experiments or manufacturing in the lab modules. Such precision is essential to the handling and repair of precision lab instruments even requiring a level of miniaturization not normally addressed in robot structures. Certainly, precision robots can also be used to make critical dimensional measurements as a means to prepare for required repair or adjustment. It would be highly desirable if a robot could repair a neighboring robot system. If any of these functions involve the human operator, precision in the slave robot would accelerate his task rate, reduce his fatique, increase his work time span, and reduce the need for his full concentration during oversight.

8. Load Capacity

Two competing criteria face the designer of robot systems for the space station. On the one hand, the robot must be generic and able to carry out a large collection of functions including those with precision even under force disturbances. On the other hand, the requirement for precision under load directly effects the overall weight of the robot which must be as light as possible. Serial robot structures which now predominate in industrial robots are least likely to meet both of these criteria. Parallel robots, similar in architecture to the Odetics walking machine, can be both low weight and able to maintain a large force with minimal deformation. Because weight is at such a premium, actuators to drive these structures must be carefully balanced in terms of power density, stiffness, weight, responsiveness, and resolution. Long experience by top motor manufacturers show that the best combination is a rare earth d.c. motor attached to a high quality harmonic drive with a reduction ratio of 180 to 1 or more. Some have recommended that direct drive motors which no gear reduction be used. Unfortunately for space station application they fail for two reasons. They will be heavier than other solutions and they will not be sufficiently stiff to control force disturbances to the system which is essential for precision operations.

Early designs for the space station suggest a 5 ft. reach robot capable of developing a 50 lb. force at the end-effector with a 1% total deflection (a deflection of 0.6 inch). Such a system could perform only the simplest of operations and would fail to be able to treat the range of operations that will be required to keep the space station available. Consequently, an integrated development effort must by undertaken to develop lightweight robots which can still maintain precision under load.

9. Reliability

Robots for the space station will have to operate in a hard vacuum, in radiation, experience thermal gradients, and be impacted by micro-meteorites. Nonetheless, these robot systems must be as reliable as possible. Failure might mean the high cost of total replacement. Or it would mean that the robot would have to be repaired by a neighboring robot system. This maintenance objective would best be met by using robots made up of modules which could easily be replaced. Redundancy in some of the hardware components (sensors, encoders, local microprocessors, etc.) can be helpful. Unfortunately, the need to be lightweight and compact makes reliability more difficult to achieve. Self monitoring software similar to that being used in advanced computers would be highly desirable. In this regard, self-calibration of the robot system after maintenance or component

replacement would be necessary to maintain the match between the control software and the robot hardware.

10. Obstacle Avoidance

Since low weight and compactness is a top space station priority, the operating environment in the lab modules will be constrained and cluttered with obstacles. Similar obstacles surrounding the space station itself will make passage uncertain. Consequently, collision avoidance technology must be part of the operating software of these robot systems. The best means available is to use the documentation of the data base to mathematically prevent collisions between the robot structure and the obstacles from occurring. This can be accomplished by putting an imaginary barrier a finite distance from the surface of the robot which when breeched by an obstacle would generate an imaginary force in the software to stop the robot from proceeding. This force could also be transmitted through the manual controller to alert the operator of the approaching physical impact. This collision avoidance technique works best with slender serial robot structures. Also, dynamic stopping procedures will have to be part of this technique. The collision problem is somewhat more difficult for the dual arm system proposed for the vehicle to be used for in-situ satellite maintenance.

11. Force Sensing

The force level experienced at the end-effector of a robot is critical to determine whether a given task is being performed properly; to determine if damage is occurring to the part being manipulated, or to be aware of excessive forces in the robot itself. Also, since low weight is important for space station robot systems, large deformations will occur. These deformations should be eliminated in order to maintain the required precision and to reduce confusion experienced by the operator of such a system. To be able to compensate for these deformations, it is necessary to have a complete parametric model of the robot structure, to calculate all link deformations due to measured forces on those links, and to develop corrections to the actuator commands to remove the unwanted deflection of the end-effector, all in real time.

Force sensing can be used to track a constrained motion as might occur when turning a crank with a robot. Force feedback to the human operator is probably the most important process parameter to assist him in carrying out complex operations. Force sensing in dual robot systems is especially important because of their high level of redundancy. Unique to the space station, it will be necessary to monitor the forces at the base (shoulder) of the robot in order to minimize negative effects (such as vibrations) on the space station structure.

12. Smoothness of Operation

Smooth operation of robot systems means that a minimal amount of dynamic shock occurs either in the command signals of the robot, at its end-effector, or within the structure of the robot itself. Dynamic shock leads to vibrations which would occur easily in space station structures because of their low weight and low internal damping. Smoothness is particularly important in the lab modules since any level of vibrations will impact negatively on either experiments or manufacturing in progress. Backlash and

stiction are shock generators and are particularly damaging in the micro-g gravity field since the structure is effectively unloaded (the equivalent of no preload to insure contact among joining parts). Backlash in actuator gear drives must be eliminated as a top priority. Stiction makes high resolution and a small minimum step difficult to achieve.

Smoothness is particularly important in precision light machining operations and to precision assembly. Lack of smoothness reduces the value of the data base as a means to accurately control the robot. Because of additive errors in serial structures, they are more sensitive to stiction and backlash. Hence, enhanced precision can be expected from t'e use of parallel structures.

13. Operational Envelope

The present Remote Manipulator System (RMS) of the shuttle has a 55 ft. reach and a level of dexterity similar to the human arm. This serial structure is ideal for low precision deployment functions and should be continuously updated for this purpose. Beyond this, smaller scale systems should be developed, preferably with common modules, both in hardware and software. These smaller systems should be either serial (for high dexterity, obstacle avoidance, reach) or parallel (for high precision, load capacity, light weight) and some should be combinations. The scales for the operational envelope might be:

RMS	60 ft.
MRMS	10 ft.
Man sized	4 ft.
Small	2 ft.
Miniature	3 inches

The number of degrees-of-freedom also must be scaled to the task:

3 DOF	handling
6 DOF	generic motion
8 DOF	high dexterity
6 large DOF 6 small DOF	cherry picker

14. Vision

Vision has the same importance to space station robotic systems as a feedback mechanism as does force sensing. Its principal function will be to enable continuous and autonomous inspection of the space station by using the reference of the data base. Machine vision excels at very small scales which makes it ideal for searching for fatugue cracks in the space station structure. Unfortunately, vision is not the ideal mechanism to sense the effects of disturbances in precision operations since it measures positional errors. It is far better to measure a higher level parameter (say force) which can then be integrated to completely eliminate the error by feedforward compensation. Nonetheless, vision is the dominant means of feedback for the operator to rapidly access the global condition of a work scene, to approximately locate obstacles, or to sense the likelihood of an unusual circumstance such as an impending accident.

Machine vision can be computer enhanced, filtered, transformed, etc. to

the benefit of the human operator. Vision technology is essential to the training function. Today it is possible to use fiber optics in the finger tips of end-effectors to make very close inspection feasible.

X. SPECIFIC PRIORITIES FOR FUTURE ROBOTICS DEVELOPMENT

This section of the paper presents an informal analysis of the near and long term development priorities for the implementation of robotics to support the operation of the space station and its various missions. The key objective of this analysis is to determine the balance of the component technologies that must be integrated to meet the mission goals. Table 1 is a collection of 14 component technologies, which when integrated into a system, would create the required functional technology. Table 2 ranks (from 1 to 10) the importance of each of the component technologies to the 5 separate missions for robotics associated with the space station:

Assembly of the Space Station
Satellite Servicing
Maintenance of the Space Station
Hazardous Manufacturing and Laboratory Experiments
Maintenance of Robots

This has been given for the near term (N) and long term (L). The results of the matrix analysis for the near term is given in Table 3 which also includes a factor associated with the near term availability of the technology (today geometry is much more available relative to its potential contribution than is intelligent control, vision, or AI). This availability factor has been used to weight the associated component technology to give a combined importance-availability indicator for the near term. Table 3 shows that geometry (10) is the most important near term technology followed closely by prime movers (9.0), graphics/CAD (8.8), etc. Table 3 should be considered as a guide to structure the balance of laboratory manpower to create near term system technology for space station robotics.

Table 4 ranks the long term importance of the component technologies (considered equally available in the long term). Not surprisingly, man-machine interface becomes a top priority for space station robotics, followed closely by end-effectors, actuator modules, sensor technology, etc. Interestingly, none of these component technologies are unimportant since the least still has a strong ranking of 6.3. This means that future robotics technology for the space station must be a balance of all technologies; i.e., the meaning of integration to meet a given mission.

Table 5 gives a listing of criteria for success of the system technology. This collection of system "quality of operation" criteria tempers the enthusiasm for a given component technology. If, for example, too much development of a given component technology is undertaken at the expense of other important component technologies, the mission effectiveness of the overall system will suffer accordingly. Each of these criteria have been ranked for each of the specific applications of robotics for the space station in Table 6 (for both the near and long term). The overall ranking of the criteria for the long term is given in Table 7. Again, as expected, the most important criteria is the generic character of the system; i.e., how many

TABLE 1: MATRIX OF COMPONENT TECHNOLOGIES FOR ROBOTIC SYSTEM

The integration of numerous technologies is one of the fundamental realities of robotics (or more generally, intelligent machines). Often significant progress in the system development will occur after a breakthrough in a component technology. Hence, except for exceptionally large research facilities, most research efforts will pursue a few component technologies and look to the manufacturer to do the system integration and development. The following 14 component technologies are given to cover the broad spectrum represented by robotics.

1. The structural geometry of the robot, its design and operation for determination of its workspace, reach, dexterity, obstacle avoidance, etc.
2. Structural dynamics of robot systems for modeling of robot dynamic and vibration phenomena for purposes of design and improved operation.
3. Prime movers are the muscles of the manipulator whose precision of operation is dependent on their response and resolution.
4. Actuator modules involves the structural integration of prime movers into modules of 1, 2 or 3 degrees of freedom which can be assembled into robotic systems.
5. End-effectors are the interface hardware and software to perform the handling, inspection, machining, etc. task of the robot; they may include special touch and force sensors.
6. Graphics/CAD of robot phenomena to enhance interactive design and optimization of robotic systems and their integration in complex manufacturing environments.
7. Sensor technology is essential to the existence of an intelligent machine so that it is aware of its own existence and process parameters associated with its operation (manufacturing, maintenance, etc.).
8. Vision is the specialized sensor capable by computer enhancement of rapidly digitizing the physical environment of the robot allowing for comprehensive planning and tactical operation.
9. Artificial intelligence structures the decision making process for multi-layered phenomena in the robot system.
10. Intelligent control involves the layered implementation of various control strategies on global and local objectives.
11. Software modules implies the compact and hardened packaging of frequently used algorithms and their specialized chip assemblies.
12. Computer architecture involves the assemblage of serial and parallel processors capable of treating multi-faceted computational tasks within the concept of real-time operation of the system.
13. Communication interfaces involves the structural distribution of operational decisions and data reduction and transfer of the sensor signals among the various components and layers of the total system.
14. Man-machine interface allows direct human communication with the intelligent robot to facilitate human augmentation in unstructured task applications as micro-surgery, nuclear reactor maintenance, etc.).

TABLE 2: ESTIMATES OF LONG TERM IMPORTANCE OF COMPONENT TECHNOLOGIES FOR SPACE STATION OPERATIONS

Estimates of Importance of Robotic Technology for Various Applications / Robotic Component Technology	Assembly of Space Station		Satellite Servicing		Maintenance of Station		Hazardous Manufacturing & Experiments		Maintenance of Robots	
	N	L	N	L	N	L	N	L	N	L
1. Geometry	4	5	3	4	4	6	6	8	3	4
2. Dynamics	6	8	3	5	3	5	5	6	3	4
3. Prime Movers	6	6	5	5	6	6	5	5	4	4
4. Actuator Modules	6	8	5	7	7	9	6	8	4	5
5. End Effectors	4	6	6	8	7	9	6	6	7	9
6. Graphics/ CAD	4	6	6	7	7	8	3	4	5	6
7. Sensor Technology	4	6	6	8	6	8	5	7	5	7
8. Vision	4	5	7	7	7	7	5	5	5	5
9. Artificial Intelligence	4	4	4	6	7	8	5	5	4	4
10. Intelligent Control	6	7	5	6	7	7	5	5	4	4
11. Software Modules	6	7	5	6	7	7	4	4	4	4
12. Computer Architect	5	6	6	6	7	9	5	7	3	5
13. Communication Interfaces	5	5	5	5	7	8	5	5	4	5
14. Man-Machine Interface	7	9	7	8	7	9	5	6	7	9

N—Near Term   L—Long Term

TABLE 3: NEAR TERM RANKING OF COMPONENT TECHNOLOGIES FOR SPACE STATION ROBOTICS

	Component Technology	Near Term Availability (Normalized)	Near Term Ranking (Normalized)
1.	Geometry	1.0	10.0
2.	Prime Movers	0.7	9.0
3.	Graphics/CAD	0.7	8.8
4.	Man-Machine Interface	0.5	8.3
5.	Sensor Technology	0.6	7.8
6.	Communication Interfaces	0.5	6.5
7.	Actuator Modules	0.4	5.6
8.	Computer Architecture	0.4	5.2
9.	End-Effectors	0.3	4.5
10.	Intelligent Control	0.3	4.2
11.	Dynamics	0.3	3.0
12.	Vision	0.2	2.8
13.	Artificial Intelligence	0.2	2.4
14.	Software Modules	0.1	1.3

TABLE 4: LONG TERM RANKING OF COMPONENT TECHNOLOGIES FOR SPACE STATION ROBOTICS

	Component Technology	Normalized Ranking
1.	Man-Machine Interface	10
2.	End-Effectors	9.3
3.	Actuator Modules	9.0
4.	Sensor Technology	8.8
5.	Computer Architecture	8.0
6.	Graphics/CAD	7.6
7.	Intelligent Control	7.1
8.	Vision	7.1
9.	Communication Interfaces	6.8
10.	Dynamics	6.8
11.	Software Modules	6.6
12.	Geometry	6.6
13.	Artificial Intelligence	6.6
14.	Prime Movers	6.3

TABLE 5: CRITERIA FOR ADVANCED ROBOTICS TECHNOLOGY

The following is a listing of 14 distinct criteria that may be used as indicators of the level of the technology available in an advanced robotic system and may be useful means to judge progress of the technology under development.

1. Multi-task capability means the number of different physical tasks that can be performed by the same robotic system.

2. Level of machine intelligence implies the level of integration of computer hardware, software, and artificial intelligence to make the system as autonomous as possible.

3. Time efficient operation implies the speed at which the robotic system performs its task relative to the human performing the task alone.

4. Unstructured task level suggests the level of numerical uncertainty of the operation that is to be performed by the robotic system.

5. Geometrical dexterity is an indicator of the motion range the end-effector can move through while performing physical tasks.

6. Portability and mobility implies the level of movement the total robotic system has relative to a stationary (fixed shoulder) manipulator.

7. Precision is an indication of the absolute precision of placement of the end-effector in world coordinates in response to simple numerical commands.

8. Load capacity clearly implies the ability of a robot to carry or resist a given load without major deformation.

9. Reliability is an indicator of the failure rate of the total robotic system.

10. Obstacle avoidance suggests the ability of the robot to avoid obstacles in its work environment.

11. Force sensing suggests the measurement of forces in the manipulator system to be evaluated by the machine intelligence to judge working forces or to compensate for manipulator deflections.

12. Smoothness of operation implies the lack of backlash and stiction or very large deformations in the manipulator system.

13. Operational envelope gives an indication of the working range provided by the robot without moving its shoulder.

14. Vision corresponds to shape recognition either by enhanced analog feedback to the human operator or by digitizing the scene and providing numerical shape recognition.

TABLE 6: ESTIMATES OF LONG TERM IMPORTANCE OF ROBOTIC CHARACTERISTICS FOR SPACE STATION OPERATIONS

Estimate of Importance of Robotic Characteristic for Various Applications / Robotic Characteristic	Assembly of Space Station N	Assembly of Space Station L	Satellite Servicing N	Satellite Servicing L	Maintenance of Station N	Maintenance of Station L	Hazardous Manufacturing & Experiments N	Hazardous Manufacturing & Experiments L	Maintenance of Robots N	Maintenance of Robots L
1. Multiple Task Capability	5	9	6	9	6	9	6	9	6	10
2. Level of Machine Intelligence	6	8	7	9	7	9	4	6	6	10
3. Time Efficient Op.	3	5	3	5	3	5	3	4	3	4
4. Unstructured Task Level	3	5	5	7	3	5	4	5	6	6
5. Geometric Dexterity	5	6	5	7	5	7	5	6	3	5
6. Portability & Mobility	8	10	7	9	7	9	5	6	4	5
7. Precision	6	7	7	9	7	8	6	8	6	9
8. Load Capacity	5	5	3	4	5	5	3	4	3	5
9. Reliability	8	8	7	9	7	7	6	8	4	5
10. Obstacle Avoidance	6	8	5	6	6	8	4	8	4	5
11. Force Sensing	6	6	5	7	6	6	4	6	6	8
12. Smoothness of Operation	8	8	5	7	6	8	3	5	3	5
13. Operational Envelope	6	8	5	7	6	8	3	6	3	3
14. Vision	6	6	5	8	7	7	3	6	4	6

N--Near Term L--Long Term

TABLE 7: RANKING OF CRITERIA FOR SUCCESS FOR SPACE STATION ROBOTICS

	Characteristic	Normalized Ranking
1.	Multiple Task	10
2.	Level of Machine Intelligence	9.1
3.	Precision	8.9
4.	Portability and Mobility	8.5
5.	Reliability	7.6
6.	Obstacle Avoidance	7.6
7.	Force Sensing	7.2
8.	Smoothness of Operation	7.2
9.	Vision	7.2
10.	Operational Envelope	6.9
11.	Geometric Dexterity	6.7
12.	Unstructured Task Level	6.1
13.	Load Capacity	5.0
14.	Time Efficient Operation	5.0

different tasks can one system perform? This criteria for multi-task capability is followed closely by level of machine intelligence, precision, portability, and mobility, etc. Generally, no one criteria is unimportant. Specifically, any one criteria could become dominant for a given application. For those who must now structure a program to produce results 15 to 20 years from now, it is clear that no component technology or no criteria for system operation can be ignored to assure success. The principal conclusion is that balance of all technologies and criteria for success is essential for the program to meet its long term mission.

XI. CONCLUSIONS AND RECOMMENDATIONS

It is believed that the NASA program for automation and robotics is potentially the most comprehensive and balanced program ever pursued for the development of robotics in the United States. This probably is because NASA has had a long term interest over at least two decades and has maintained some in-house development groups for this purpose. Without doubt their appreciation for the importance of the man-machine interface is unique and more balanced than found in any other agency program. Also, NASA has been properly tasked to provide an enhanced tech base to respond to our productivity problems associated with the weak U.S. trade stance in value-added goods (a deficit now of not less than $100 billion/year). Generally, this mission would be best associated with the Department of Commerce, but it has little history of managing a major technology based development program. Because of NASA's high regard for integrated technology and its need to be mission oriented, its successful pursuit of the technology for robotics is more likely.

The principal operation (80%) of the space station will be satellite servicing as much as possible in-situ. The service module will always be weight limited with a finite number of on-board replacement modules for repair in terms of a very generic robotics capability to perform complex precision tasks some of which will require dual robot arms. The manufacturing and experimental lab modules are not a top priority of the space station although important. These lab modules must be compact and lightweight and as a consequence they will be cluttered and unattractive for robot operation. Snake type robots made up of modules which can be easily added or removed give the system the ability to reconfigure itself to best meet functional needs within these lab modules. Because of the wide range of needs represented by the space station, it is clear that major improvements in robotics technology must occur in order to make the station both feasible and economical. This need for robotics results from the fact that EVA for astronauts is either too time consuming, unsafe, or unable to meet the needs of complex precision operations.

The most serious issue for the development of robotics for the space station is associated with the penalty of weight. For example, the 5,500 lb Cincinnati Milacron T3-776 industrial robot deforms 0.1 to 0.2 inch under its payload of 150#. This machine is much more robust than most industrial robots. The technology represents the second generation of robots where disturbances from the process (forces on the end-effector) cause major uncorrected deformations because the system operates passively in an open loop mode. If we now reduce the robust nature of these machines so that they weigh

less (say 550 lbs) by a factor of 10, then the ability of the present system technology to perform anything but the simplest functions will be lost. This means that excess deformations will confuse either the human operator or the computer driving the system from a database.

The demonstrated technological limitation is associated with what might be called "intelligent control" or real time dynamic modeling with adaptive control to make the system electronically rigid as is done for our fly-by-wire aircraft. Without this technology, many of the critical precision functions:

light machining
welding and forming
disassembly and assembly

found throughout the application spectrum of the space station mission will be unmet. Yet, the technology is feasible. What must happen is to structure a development program today to treat:

metrology of robots
modularity of robots
mechanical architecture
real time computation software
modern control technology for robots
adaptive control
etc.

Most of these topics represent a marriage of the most modern electrical and mechanical technologies. This need for marriage and balance must be a guiding principal if the technology is to be adequate to meet the missions associated with the space station.

Some conclusions and recommendations can be made relative to the development program for space station robotics. These are:

1. Because extensive experience has been gained to evaluate the use of robotics in nuclear reactor maintenance, much can be learned from that experience to predict the parameters of the needed technology for maintenance operations on the space station. This is especially desirable since the space station is essentially a one-of-a-kind facility.

2. The DOD logistics maintenance and parts on demand activity (jet engine repair, airframe repair, etc.) are similar to many of the maintenance operations that will be necessary for the space station. Hence, these two communities should establish joint programs in order to move the technology forward more economically.

3. The Secure Automated Fuel Engineering facility (SAFE) involves a range of operations which are very similar to those expected in the lab modules. Again, joint programs can move the technology forward more quickly and at a lower cost.

4. The preceding analysis shows that all component technologies are relatively important to the space station operation as are all the various criteria to measure progress in the system technology. Hence,

the research and development program must be balanced with no one component technology being given too much or too little attention.

5. Three levels of system technologies that are very immature in robotics must be vigorously brought forward for use in the space station:

   Snake robots — dexterity
   Dual robots -- constrained extra DOF
   Walking robots — balanced extra DOF

6. The future of robotics (that needed in the space station) is dependent on the balance between an enhanced understanding of deterministic analytics (based on physical phenomena) and underterministic decision making (based on AI principles).

7. Since the technology for space station robotics is very far reaching and since it must be highly generic, it is recommended that NASA pursue the development of an extensive collection of modules for robots (i.e., 1,2, and 3 DOF modules, sensors, end-effectors, etc.) from which any future robot could be rapidly assembled. This would be the most economical means of near term development making system development much more rapid once the space station needs are more fully documented.

8. It is recommended that direct drive motors be considered but that a major effort be established to pursue the development of special lightweight rare earth D.C. motors connected to the uniquely appropriate harmonic drive, as a class of powerful lightweight actuators for space station robot systems.

9. The architecture of both the computer hardware and software must match the architecture of the robot geometry and control software. This marriage is critical to the ability of the computer to actually drive a "known" robotic structure.

10. The computer must drive the actual robot structure "as is" not "as built". The difference will be a significant variation in the physical parameters from those used to construct the computer software. This means that each robot must be recalibrated each time it is reconfigured, adjusted, or maintained. This is a question of a very high order.

11. NASA should establish standards for communication interfaces very early in the development of the space station.

12. Man will be required to enter in (intervene) at a higher technological level as the techology progresses. This means that the man-machine interface will become more important—not less, as many suggest. The recommendation is to concentrate on a universal force feedback manual controller capable of driving any robot in real time. -

13. A dominant system criteria for space station robotics is that they must perform a very wide range of physical tasks (in scale, precision, dexterity, speed, disturbances, etc.) in order to minimize the number of robots needed (reduced weight) and the likelihood that emergencies can be met by any available robot.

14. All documentation in the data base of the space station must be continuously updated to represent the "as is" status at all times. Otherwise, the burden on the crew to supervise the robot systems would be excessive.

15. Since the technology for space station robotics will be technologically different from that of existing industrial robots and since the need to achieve near optimal results is an imperative, the methodology of design for space station robots will have to be based on sound engineering principals and the best existing and future CAD technology, some of which will have to established for this purpose by NASA.

16. An important facet of the space station program statement provided by congress is that NASA is to transfer the technology it develops from this program to enhance productivity in civil sector manufacturing. Hence, it will be necessary for NASA to evaluate the weaknesses plaguing the nation in its ever increasing trade deficit from high value added goods so that it can properly prioritize this transfer.

In several supporting documents, the role of the universities is considered to be unresponsive to the mission goals of a development program of this magnitude. If the goal is also to create infrastructure to create a national response, then young people must be involved. To do so means to change the way we structure the university program element, supporting those university centers that present a team approach with industry to meet development objectives in a timely fashion. This can be achieved, but it must be an early program priority carefully worked out by the decision makers of NASA, the affected universities, and the cooperating industries.

The following recommended automation and robotics technology program cost summary* appears to be well balanced and proportioned to the need described in this assessment although the funding for the prototype demonstrations appears somewhat low. Nonetheless, some reports and public presentations do not appear to reflect these proportions leaning heavily in favor of one or two component technologies. This bias may be a result of the mistaken belief that one or two technologies will fully meet a given mission since the remainder can be bought off the shelf when the need arises. Everything this assessment has encountered argues very strongly for the pursuit of all component technologies and system characteristics in proportion to a detailed analysis of those needs. A structure for that analysis is given in this assessment and some numerical priorities have been suggested although a more careful analysis is recommended.

* See Ref. 12.

277

Cost Summary ($M/year)

	Research	Demos
NASA leads		
Man-Machine Interface/Robotics	51	10
Information Management	26	4
NASA Applies Leverage		
Man-Machine Interface/Robotics	25	4
Information Management	31	7
NASA Exploits (Buys)		
Man-Machine Interface/Robotics	1	1
Information Management	4	1
Infrastructure (Technical)	23	5
	160/yr.	30/yr.

This assessment has stressed the need for balance in the development of technology for automation and robotics for the space station. The first reality is that the space station is a one-of-a-kind effort and that its operation is at least as complex as that of a nuclear reactor. The history of these reactors is that they are available for only 75% of the time. This lack of availability in the space station would be devastating to its usefulness to the military and especially to the strategic defense initiative. The best opportunity to improve this availability (say to 95%) would be through the full implementation of an advanced robotics technology to perform service and maintenance especially under emergencies or attack. Unless the availability question is dealt with directly, it is likely that the total investment will be poorly used. The history of this type of system (nuclear reactors, ocean floor activity, etc.) clearly suggests that care must be taken in managing the program and, in this case, it means balance and a due regard to the required contribution needed from all technologies.

REFERENCES

1. "NASA Space Station Automation: AI-Based Technology Review", Oscar Firschein (PI) et. al., Report for NASA Contract No. NAS2-11864, April 1, 1985.

2. "Advancing Automation and Robotics Technology for the Space Station and for the U.S. Economy", NASA Technical Memorandum 87566, Volume 1, Executive Overview, Volume II, A Technical Report, March, 1985.

3. "NASA Automation and Robotics", May 13-17, 1985 Workshop, Johnson Space Center, Vols. I and II.

4. "Space Station Automation Study", Martin Marietta Report for Contract No. NAS8-35042, Vols. I and II, November, 1984.

5. "Automation Study for Space Station Subsystems...", Hughes Aircraft Co.'s Report for NASA Contract No. 82-14F, Project Leader, J. T. Yonemoto.

6. "Space Station Automation and Robotics Study, Operator-Systems Interface", Boeing Co. Report for NASA Contract No. D483-10027-1, November, 1984.

7. "Space Station Automation Study, Automation Requirements...", General Electric Co. Report for NASA Contract No. NAS5-25182, July, 1984.

8. "Space Station Automation Study, Satellite Servicing", TRW Report for NASA Contract No. NAS8-35081, Project Leader H. Meissinger, November 30, 1984.

9. Testimony on Space Station Automation and Robotics by Aaron Cohen to Subcommittee on Science, Technology and Space, United States Senate, March 8, 1985.

10. "Next Generation of Technology for Robotics", D. Tesar, Presented to the NASA Symposium for Space Station Automation and Robotics, September 4-6, 1985, Washington, D.C.

11. "Assessment for the Design and Implementation of Robotics to the Secure Automated Fuel Fabrication Plant", D. Tesar, et. al., University of Florida Report, October 1, 1983.

12. "Automation and Robotics Panel Report", Program Director, D. R. Criswell, California Space Institute, University of California, San Diego, NASA Grant NAGW 629, February 25, 1985.

APPENDIX

LIST OF RESEARCH PROJECTS

FOR THE

PROPOSED CENTER FOR ROBOTICS, MANUFACTURING AND LOGISTICS

BY

DELBERT TESAR

1. Metrology of Robots
Sponsor: NSF and the Department of Mechanical Engineering.

In order to create the dynamic model of a robot structure as the basis for its control, it is necessary to numerically identify all system parameters such as link dimensions, link masses, structural spring rates, prime mover control parameters, etc. Only one robot is known to have its operating software adjusted to reflect the "as built" link dimensions. All other robots operate on their "designed" link parameters making them inaccurate under direct computer control. The research project will establish a laboratory for robot metrology in terms of advanced vision and analytical modeling tools to semi-automatically identify all significant parameters of existing industrial robots.

2. Optimal Robot Design
Sponsor: Department of Mechanical Engineering

There are a large number (108) of individual parameters in a 6 Degree-of-Freedom serial robot structure (18-geometry, 42-mass, 36-deformation, 18-actuator control). Even for this simple geometrical structure, the design problem is truly massive. For example, what is the optimal actuator load capacity distribution along the arm to ensure maximum payload at the end-effector? Or what is the optimal distribution of actuator stiffness values to ensure the most precise operation while performing unique light machining operations? Initial success in the use of optimization techniques to the multi-parameter multi-criteria problem associated with robotics has led to improved distribution of these actuator parameters. This computationally intensive effort must be expanded to include all system parameters for criteria applied throughout the work volume of the robot.

3. Robot Architecture
Sponsor: The Department of Mechanical Engineering

Future robots will be composed of easily scaled structural modules (shoulders, wrists, micro-manipulators, specialized end-effectors, mixed large and small control structures, etc.) to provide finite packages of proven technology to be rapidly assembled into generic intelligent machines. This type of modularity allows the rapid diffusion of new technology without disturbing the more slowly evolving system architecture at much lower cost. Maintainability and rapid up-dates of obsolete modular units would become much more feasible. The research team has already conceptualized several modules and is working on their design, scaling, and interface requirements in the same manner as computer chips and boards are now used to assemble modern computer systems.

4. Operational Software
Sponsor: ONR and the Department of Mechanical Engineering

The fundamental immediate need to be satisfied for control of robotic structures is the complete real time description (less than 30 msec) of the dynamic model of the system. Thus far, the research team has been able to calculate the model of a completely general 15 link serial structure (15 DOF) in 30 msec. This system is composed of rigid links. Future descriptions will include n dominant deformation modes in selected links as n s[illegible] quasi DOF within the same format for the rigid N DOF structure. Symbolic programming is now being applied to these complex analytical formulations in order to form the basis for a generic operational language of on-line programming and disturbance rejection. This is believed to be the first major effort to create a completely general language structure to treat the dynamic description of robotic systems in terms of generalized matrices (Jacobian, Mass, Coriolis mass, etc.) and to formally organize the development of the controlling equations. Initial results were obtained to determine the best computational resource allocation for various descriptive terms in the dynamic formulation. This work is now being carried out on selected array processors and will also be transferred to the super computer in the near future.

5. Adaptive Control
Sponsor: ONR and the Department of Mechanical Engineering

This adaptive control scheme adjusts the control laws of the prime mover system to reflect the real time condition of the full non-linear and coupled nature of the mass and external load (disturbances) effects on the stability and precision of the system. The claim for global stability is based on a Liapunov analysis. Initial results are obtained on the effects of computational sampling rates and the associated assurance of stability. Preliminary work on criteria for precision of the system's end-effector motion is also being pursued.

6. Modeling of Complex Robot Structures
Sponsor: ONR and the Manufacturing Systems Engineering Program

As robotic structures become more general, devices such as walking machines (4 and 6 legs with several inputs each), dual arm robots (a total of 12 inputs) must be created. Their generalized dynamic model formulation (both for serial and parallel structures) is developed in terms of influence coefficients in order to allow the designer complete freedom to locate his prime movers in ideal positions in the structure. In addition, formulations are being developed for a third order description of the dynamic properties of the system as well as a means to mathematically transfer any or all of the prime movers to any location in the structure.

7. Vibration Analysis of Robot Structures
Sponsor: ONR

The deformations in the serial structure of robots is obtained by representing the six modes of deflection of each link as pseudo joints (6N) which can be added to the N prime mover joints by means of the influence-coefficient method. A very complete model formulation is being establishea to account for end-effector deflection due either to applied external loads or inertia loads. Also, a method is given to compensate for these deflections in order to eliminate their effects and maintain the desired precision of operation even under disturbances. A lumped mass and deformation model described in terms of influence coefficients with pseudo inputs at the principal deflections is used to model the vibratory motion and to predict the frequencies of the lower modes. In addition, modern modal analysis equipment is being used to analyze a Cincinnati Milacron T3-776 robot to identify local stiffness and mass parameters and to experimentally verify the vibration model and frequency predictions of the lower modes.

8. Man-Machine Interface
Sponsor: Manufacturing Systems Engineering Program

As the technology becomes more complex, a greater need (not less) will develop for a balanced control (or intervention) by man and machine. This will require a much higher level of machine intelligence to obtain the full benefit of the technology for man. Robotic systems used in unstructured task environments such as nuclear reactor maintenance require the full integration of the human operator's judgement and decision making capability. This effort has established the design criteria for a kinesthetic force feedback manual controller of one extra degree of freedom as an assemblage of structural modules at the wrist and shoulder. The goal is to enhance the flow of information to and from the operator in real time and to use supervisory techniques to eliminate gross errors, filter jitters, and perfect the global input commands from the operator.

9. Machining Robot
Sponsor: ONR and the Manufacturing Systems Engineering Program

The heart of the factory of the future will require inexpensive generic robots to perform precision light machining operations by direct computer control in order to have a maximum value-added benefit and response to the individual consumer. This requires a complete dynamic and vibration model implemented with feed-forward compensation in real time to make the system electronically rigid in order to maintain the required level of precision without the support provided today by jigs and fixtures. Elimination of the jigs would dramatically reduce start-up costs and the length of the learning curve now experienced by batch mode operations such as found in airframe manufacture and would also allow direct real time process monitoring by the factory data base to ensure quality control.

10. Six DOF Micro-Manipulator
Sponsor: Navy Mantech Program

A 6-DOF parallel structure small motion device (±0.1 inch, ±2 degrees) has been designed as a module weighing 20 pounds and an overall size of about a 7" cylinder 5" high. This module would be placed between the end-plate of the robot and the end-effector to make very small motion corrections much more rapidly than is feasible by the large actuator control system normally found in robots. Influence coefficient analysis is being used to create a dynamic model of this device and to establish design criteria for its most effective operation and control.

LJI-R-85-351

STRATEGY FOR COMPLEX ORGANIZATION MODELING, PLANNING AND EXPERIMENT (SCOMPLEX)

BY

BRUCE J. WEST
DIVISION OF APPLIED NONLINEAR PROBLEMS
LA JOLLA INSTITUTE
P. O. BOX 1434
LA JOLLA, CA 92038
(619) 587-6000

JULY 1986

THIS RESEARCH WAS SPONSORED BY THE
DEFENSE ADVANCED RESEARCH PROJECTS AGENCY
UNDER ARPA ORDER NO.: 3710
CONTRACT NO.: MDA903-82-C-0187

280

TABLE OF CONTENTS

351

EXECUTIVE SUMMARY

This report outlines and motivates a mechanism for managing complex, large scale, scientific and technological research programs. This strategy (S) for complex (C) organizational (O) modelling (M), planning (PL) and experimenting (EX), (SCOMPLEX) is discussed with reference to examples drawn from physical processes relevant to DARPA programs. Drawing from our experience in the analysis of complex systems, we argue that: if the systems consist of a number of only weakly interacting components, (example, ASW) then the traditional managerial schemes would appear to be adequate. If however, the system consists of a large number of strongly interacting or interdependent components (example, wave propagation in random media), then a new managerial scheme is proposed. This new scheme has the following elements:

1) A principal investigator as the *Program Director* and who has the final decision making responsibility in all program areas.

2) An *Overseer Committee* consisting of the senior scientists in the program which will oversee the allocation of funds, personnel etc. as well as the overall research direction of the program.

3) An *External Scientific Advisory Panel* (ESAP), whose members are chosen from the academic, industrial and government scientific committees to provide the best available advice and guidance from the outside experts in the field of programmatic interest.

4) A *Technical Transfer Panel* (TTP), whose member will be chosen from governmental and industrial settings to provide the best available advice and guidance in the technological implications of the scientific research, particularly for the services. It is suggested that the chair of this committee be the DARPA contract monitor for the program,

5) These four elements are duplicated at the level of the research conducted by each of the *senior scientist* in the program. In this way they each become the Director of their individual project for which there is an Overseer Committee, an ESAP and a TTP. The managerial structure is thus seen to be self-similar and the number of self-similar levels depends on the degree of complexity of the program.

6) At each level of the proposed hierarchy the four elements are used in a self-assessment mode to determine if the research goals of the program are being realized and if not what is required to realize their goal. They monitor, critique and guide the scientific research in a coordinated manner.

287

1. INTRODUCTION

A strategy is formulated for the management of complex technological programs of interest to DARPA and the DOD. Large scale programs such as the *Strategic Defense Initiative* (SDI) bring into focus the need for a comprehensive management approach that can identify and disseminate methodologies that have been developed and verified within one context, to other contexts wherein the complexity of the problem has proven to be insurmountable.[1-4] This strategy has been adopted in the development and application of certain technologies, but has heretofore not been used in the refinement of formal research techniques required to solve classes of applied problems. For brevity, let us refer to such an undertaking as a strategy (S) for complex (C) organization (O) modeling (M), planning (PL) and experiment (EX) and denote it by the acronym SCOMPLEX. It is often easier to present the scope of a program such as SCOMPLEX in general terms, since to achieve generality, one is forced to blur-out the details. The vagueness that results when the details are missing often leaves one with a feeling of dissatisfaction. To circumvent this pitfall, we have elected to present a number of examples, on whose importance we can all agree, and through the development of these examples, expose all those ingredients that such a program strategy must blend.

In the following sections we argue that complex systems fall into categories, those that can be segmented into isolated (or weakly interacting) units (category 1) and those that cannot (category 2). Most research programs are built under the assumption of category 1. We will discuss ASW as the prototype of such a research program. This is then contrasted with how one develops an understanding of a particular physical mechanism or process that can not be further decomposed. We discuss wave propagation in random media as the paradigm of this latter type of system.

1.1 Complex Systems and Predictability

Whether one is discussing the micro-mechanics of crack propagation leading to the disintegration of a turbine blade during operation, the failure modes of the integrated circuits in a super-computer, the physical properties of disordered materials, or even the generation of decision policies in conflict scenarios, complex behavior emerges out of a confluence of relatively simple parts. The individual parts either are understood, or can be understood using traditional research modes, but the comprehension of the overall system complexity is quite another matter. The strategy of partitioning a complex process into its constituent elements, studying the separate

WORLD VIEWS

DETERMINISTIC	PROBABILISITC (CHAOTIC)
• SIMPLE SYSTEMS	• COMPLEX SYSTEMS
- Single component	- Many components
- Low velocities	- High velocity
- Low energy	- High energy
• STABLE DYNAMICS	• UNSTABLE DYNAMICS
- Linear	- Nonlinear
- Laminar fluid flow	- Turbulence
• STABLE STRUCTURES	• UNSTABLE STRUCTURES
- Attractors	- Strange attractors
- Solitons	- Dissipative structures
• CHARACTERISTIC SCALES	• NO CHARACTERISTIC SCALES
• SYSTEM IS DECOMPOSABLE	• SYSTEM IS NOT DECOMPOSABLE

Figure 1.

ASW

SIGNAL SOURCES

- DIRECT MODULATION OF SURFACE WAVES BY VEHICLE GENERATED CURRENTS
- VORTEX TRAIL
- WAVE GENERATED BY HULL DISPLACEMENT
- CURRENT ASSOCIATED WITH THE HULL
- TEMPERATURE DIFFERENTIAL

NOISE SOURCES

- WIND GENERATED SURFACE WAVE SPECTRUM
- WIND GENERATED SURFACE CURRENTS
- WAVE-WAVE TRANSFER OF ENERGY
- FLUCTUATIONS IN THE WIND/WAVE FIELDS
- MODULATION PATTERNS PRODUCED BY NATURAL SOURCES

SENSORS

- RADAR BACKSCATTER FROM SEA SURFACE
- LIGHT BACKSCATTER FROM SEA SURFACE
- PASSIVE IR DETECTORS
- SONAR DETECTORS

DATA PROCESSING

- MATCHED FILTERING
- FALSE ALARM PROBABILITIES

Figure 2.

parts, and then reassembling the whole, has worked for the most part for nearly three hundred years. The approach has not been uniformly successful, however. Exceptions immediately come to mind when one thinks of applications, e.g., turbulent fluid flow, high-stress breakdown of materials, intense laser radiation, etc. Each of these examples exposes a limitation of this traditional approach when the system being studied is outside the linear regime, i.e., strong interactions among fluid elements in turbulence, long-range interactions in high-stress materials, intensity dependent index of refraction in intense laser radiation, and so on. Physical processes are not the only ones where the limitations of the traditional methods are becoming evident; other examples include living systems, formal organization, finite state machines and the majority of physical systems used in high-tech programs. The typical problems that must be addressed by scientists and program managers alike include the description of the dynamical and structural properties of complex systems, their adaptability and self-organizing capabilities and finally their fault tolerance. These properties may be summarized into the single concept of the predictability of the system properties.

In Figure 1 we depict the dichotomy of the above two world views. The first one attempts to model processes as simple deterministic systems that are stable and which can be fitted together to construct a large system of weakly interacting elements each with its own characteristic scale. Even instabilities in such systems are satisfactory described by linear stability analysis. The second view is that of complex systems where the components are viewed as strongly nonlinearly interacting and therefore cannot be separated one from the other. Such systems generate instabilities and often require a probabilistic rather then a deterministic description of its dynamics. This separation of views is over simplified, but is still useful from the point of view of a program manager: to whit, if a complex systems falls into category 1 then a program may be developed in which the research on the separate components constitute the program elements, whereas if the complex systems falls into category 2 then the problem must be addressed in its entirety. A new program management scheme is required for category 2 problems. How one can determine into which of these categories a particular program belongs in discussed in Section 3.1 and how programs in category 2 can be managed is discussed in Section 3.2.

1.2 ASW as a Complex System

We mentioned that ASW would serve as the prototype for the complex systems of category 1. In Figure 2 we present four different aspects of the ASW detection problem; 1) the candidate signatures of the vehicle that serves as a signal, 2) the possible source of noise in the geophysical

WAVE PROPAGATION IN RANDOM MEDIA

WAVE PROPAGATION

- SOUND IN THE OCEAN
- LIGHT THROUGH THE ATMOSPHERE
- RADAR THROUGH THE IONOSPHERE
- IR IN THE OCEAN AND ATMOSPHERE

WAVE SCATTERING

- RADAR FROM SEA SURFACE
- SOUND FROM UNDERSIDE OF SEA SURFACE

DISRUPTION OF WAVE

- AMPLITUDE FLUCTUATIONS
- PHASE FLUCTUATIONS
- INTENSITY FLUCTUATIONS

Figure 3.

environment, 3) various sensors that respond to the different candidate signatures and 4) the data processing techniques. It cannot be over- emphasized that because the sources of noise given in Figure 2 also act in part to *carry* the vehicle signature, that the detection is a statistical question. Thus the end result of this program is an algorithm for a probable detection that carries with it a possibility of being a false alarm. The likelihood of a detection being a false alarm can be reduced in a number of ways: 1) increased knowledge of and better models for the noise sources; 2) better models of the interaction of directly generated vehicle characteristics with the dynamic geophysical environment; 3) better understanding of sensor wave scattering and/or radiation from complex (modulated-turbulent) water wave fields and 4) being able to identify different noises in the data and to process them separately. We note that the statistical aspect of this problem arises because one of the component elements (the noise) is stochastic.[5] This is distinct from the way in which fluctuations arise in category 2 processes, the strongly interacting components of these latter systems generate their own noise even when the system elements are deterministic.[4]

As is clear from Figure 2 the aspects of the ASW problem are decomposable, at least in-so-far as they are readily distinguishable as questions relating to signal processes, noise processes or data processing. This partitioning may well be illusory, however. We will return to this point subsequently.

1.3 Wave Propagation in Random Media as a Complex Process

Let us now turn to the category 2 processes. Important examples of such processes arise from the effects of medium fluctuations on several types of wave propagation, including radio waves through plasma, light through the atmosphere, and sound through the ocean (cf. Fig. 3).[6,7] In each of these cases, the traditional approach is to obtain predictions of the correlation properties of the transmitted wave based on specific models of the media, and to compare these predictions with experimental data. These correlation properties often include the frequency spectra of phase, amplitude, intensity, and coherence function for a monochromatic wave; pulse spreading; and the associated communications channel characteristics such as the time-bandwidth product. Special attention is often given to the probability density function of intensity and its associated moments as in the study recently completed for DARPA at the Center for Studies of Nonlinear Dynamics.[7] The emphasis in that study was placed on the intercomparison of results from media of quite different physical characteristics; the purpose being to understand what results are common to all media, and to identify the relevant medium characteristics that affect results that differ between media.

One often discusses this problem as if the wave were separable from the medium through which it is propagating, or only slightly perturbed by the surface from which it is scattered.[6] When the medium is random and/or the scattering surface is rough, however, the wave problem loses its linear character in that the fluctuations enter through the "index of refraction" which is a state-dependent (nonlinear) noise. It is now understood that the traditional perturbation theory is not sufficient to analyze the properties of systems with such multiplicative noise.[7] It is known for example that even simple systems with such state-dependent noise can be "energetically unstable" and lead to diverging central moments.[27] Thus one must either solve the problem *non-perturbatively* or else apply techniques that can systematically sum selected infinite series of perturbations (renormalized perturbation techniques) in such a way as to quench the instability.

Note that in the ASW problem, the sensor wave, such as the EM wave scattering from the sea surface is usually treated separately from the scatterers. The modulation of the scattered wave by the surface wave is treated perturbatively, i.e., is a category 1 effect. This cannot be done for waves propagating in a random media because of the possible divergence of low order moments. Thus even though these two problems have a great deal in common the first one is treated as category 1, while the second is treated as category 2. In this regard it appears that category 2 systems have to be viewed as a single phenomenon or process and hopes to further reduce such systems into simpler constituent elements must be abandoned.

1.4 Deterministic Fluctuation

The general ASW problem and the description of wave propagation through a random medium each has a stochastic element. The randomness arises from the complexity of the geophysical environment and the indeterminacy of the flow field predictions.[1,2] Weather and ocean wave forecasting are natural contexts in which to address the predictability problem. Therein, this problem is quite practical, bearing upon the possible improvement of forecasting and also predicting the likely range of errors in any forecast. Weather forecasting has been the driving force behind many predictability studies; nonetheless it would be a mistake to identify predictability simply as a problem in this limited context. Rather, predictability is a fundamental theoretical and computational issue in the analysis of nonlinear equations such as in fluid dynamics[1,2] or low-order dynamical equations from other areas of physics.[3,4]

Much of classical dynamics is divided between deterministic and statistical points of view. The epitome of the former view were particle orbits in analytical mechanics while the latter view

was sometimes associated with turbulence in fluid mechanics.[3] These two views were long thought to be distinct. More recent work has established that chaotic behavior can result from deterministic systems with as few as three degrees of freedom.[4] Thus, the point of view of a segment of the scientific community has shifted more towards the perspective that purely deterministic behavior is an illusion and only distributions, albeit very narrow ones in some cases, have physical significance.[1]

Despite practical concerns as well as fundamental theoretical interest, the general area of predictability studies has not become well established within the body of fluid dynamics. Although some recent textbooks have included brief discussion of predictability issues,[1,2] the literature of predictability consists of isolated research articles scattered through a number of journals. Thus a DARPA research program which focuses on a number of important applications that require the resolution of a subset of these questions for the realization of the application would be desirable. A new kind of program coordination would be necessary to achieve this objective.

2. SYSTEMS COMPLEXITY, APPLIED PROBLEMS AND ASW

We emphasize here that traditional physics, chemistry and engineering are filled with simple laws such as momentum conservation and energy conservation, Ohm's law, Gaussian statistics, etc. Everywhere exponents are apparently integers and quantities are well defined. However, when one studies actual systems which are sufficiently complex, these simple relations may not be in evidence. Integer exponents can usually be traced back to the analytic behavior of an appropriate function which can be expanded in a Taylor series. Non-integer exponents imply the presence of singularities and the breakdown of a Taylor series due to the divergence of a coefficient. These multiple scale structures are manifest through the existence of power-law distributions which characterize the system.

In the past decade long-tailed distributions have appeared with increasing frequency in the investigation of such diverse phenomenon as turbulence,[8-11] extreme properties of stochastic systems (false alarms and failures),[12] wave propagation through random media (in the ionosphere,[6,7] in the atmosphere[6,7] and under the sea[13]), the statistics of non-linear wave-wave interactions,[5] the detection of clustered events near a critical point,[14] the properties of polymers,[15] and so on. Many other applications of these long-tailed distributions and their relation to *fractals* has been discussed in a recent conference[15] jointly hosted by the La Jolla Institute, the National Bureau of Standards, the Office of Naval Research, IBM, GE, SOHIO and the University of Maryland.

If the central moments of an appropriate observable with respect to one of these long-tailed distributions diverges, then no linear scale exists by which to gauge measurements and structure arises on all scales. The concept of self-similarity, non-differentiability, and also non-integer exponents all accompany the divergence of such low-order moments. It has recently been found that simple mechanical models can be constructed which exhibit these properties. Under a completed DARPA contract on the nonlinear properties of materials, members of the La Jolla Institute developed random walk examples where the above characteristics appear simply and naturally. These random processes have an inherent self-similar (fractal) scaling in space, time, frequency or other appropriate variable. They can be used to model complex systems of interest which exhibit features spanning many decades of scale.[16-22]

Data bases can be divided into those that capture the time-independent structural aspects of the system or those that characterize the evolving features. As shown in Fig. 4 both these types of data can have self-similarity or scaling properties. The dynamic (time-dependent) data are often in the form of a spectral decomposition of a time series. Such series display intermittent activity

FRACTALS AND SCALING

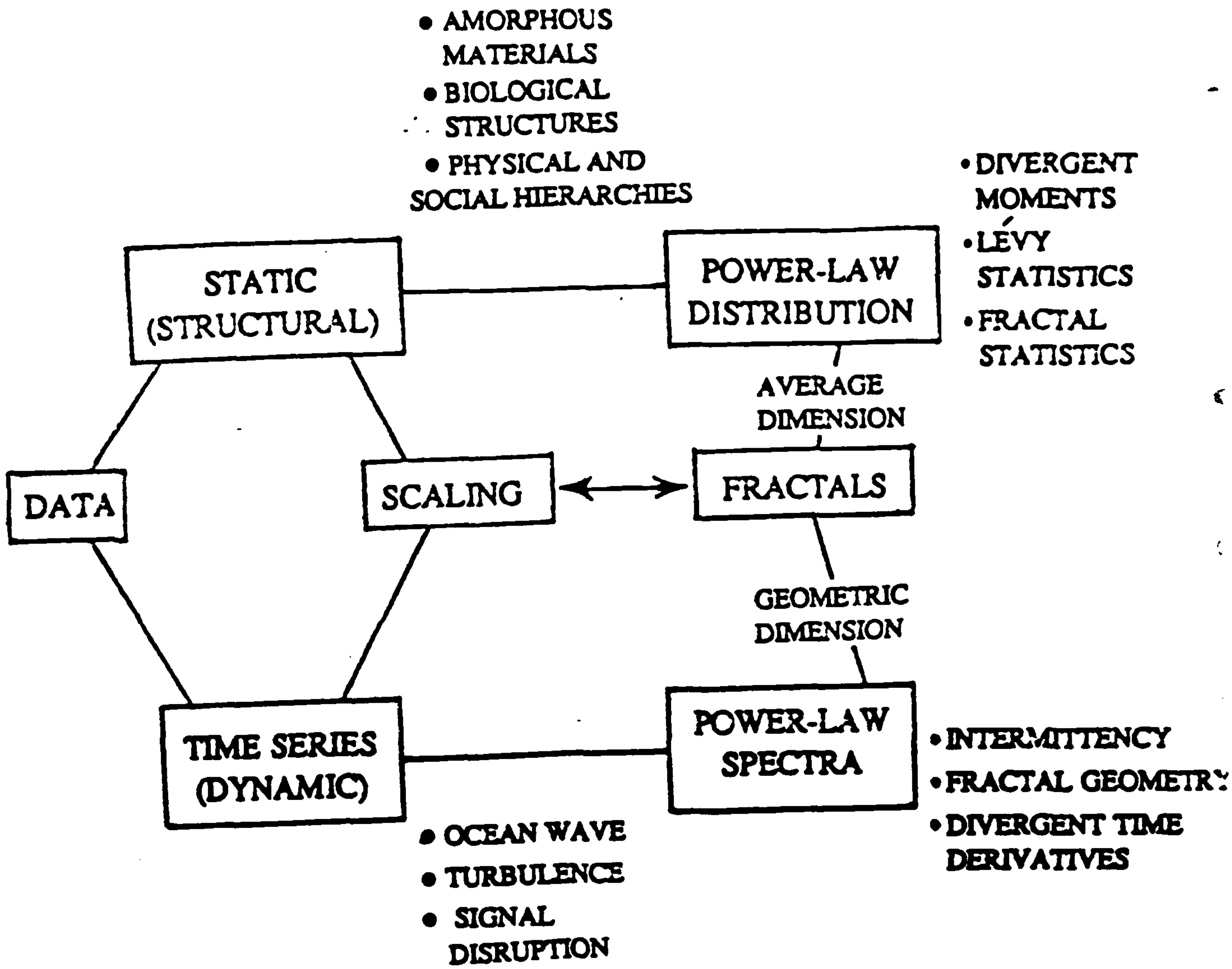


Figure 4.

when the system being probed is sufficiently complex, eg., turbulence, ocean waves, etc., and the corresponding spectra are inverse power law in nature. The power-law index can be related to the geometric notion of a fractal dimension. The data for the static system can often be expressed in terms of a distribution of the scales present in the structure, eg. amorphous materials, critical phenomena, etc. If the distribution is a power law (or Levy which has an asymptotic power-law form) then the index of the power-law distribution can be related to the statistical notion of an average fractal.[4]

To pursue our example of the ASW problem let us look in more detail at three of the sources of noise; these being the ocean water wave field, turbulence in the water and turbulence in the wind and how they affect the sensor signal. Each of these share certain of the fractal properties referred to above.

2.1 Ocean Surface Waves and Internal Waves

A turbulent wind blowing over a water surface generates a broad spectrum of water waves. The hydrodynamic equations of motion are nonlinear, so that the surface wave field is described by a system of nonlinearly coupled mode rate equations driven by a fluctuating force. This wave field is, of course, not isolated, but interacts with surface currents generated by the wind, internal waves, and other energy sources such as submarines. The richness of this problem has eluded the development of a faithful mathematical model with the capability of solving the full dynamic equations analytically.[5] However, certain model descriptions of the evolution of the isolated surface-wave field have met with some success as have models of the interaction between surface and internal waves.[5,23] It is very important to understand the role played by statistics in the evolution of the water wave field, both interactively with the environment and in isolation. A "complete" description would require a Monte Carlo simulation of a broad band spectral process (surface waves) modulated by a somewhat narrow-band process (surface currents) and driven by a turbulent wind field.

The interaction between the wind-generated surface-wave spectrum and the surface current generated by internal waves has been proposed as the physical mechanism giving rise to observed ocean surface modulation patterns as well as one of the possible sources of energy for the internal wave field[23] (cf. Fig. 5). The mechanism that produces *surface slicks* has not been unambiguously determined, i.e., whether it is a direct interaction of surface waves with surface currents or the effect of concentrating naturally occurring organic films on the sea surface by the surface

OCEAN WAVES

SURFACE WAVES

- Generation by Fluctuating Wind (linear?)
- Hamiltonian System (without wind)
- Wave-wave Energy Transfer (nonlinear)
- "Boltzmann" Equation Describes Evolution

INTERNAL WAVES

- Dominant Generation Mechanism is Unknown
- Larger Scale Than Surface Waves
- Mask Signature of Undersea Vehicle

SURFACE WAVE / INTERNAL WAVE COUPLING

- Direct Modulation by Surface Current
- Indirect Modulation via Wave-wave Interactions
- Pattern Modification by Surface Films

Figure 5.

currents, or both. In any event the identification of these surface modulation patterns is of central importance in ASW, since an under-sea vehicle can generate surface currents in a number of ways including the generation of a vortex trail. Estimates of the magnitude of the interaction between surface waves and currents have been made based on various combinations of linear perturbation theory and random phase approximations. Although some of the estimates have been successful in describing laboratory experiments and certain controlled field experiments, there does not exist to date a complete theory which is applicable to the open ocean environment.[5] Hamiltonian equations of motion provide the most systematic description of this problem;[23] it also is expected to be important to explicitly include a stochastic phase in the dynamics. This approach can describe the stochastic phenomena in the wave field while at the same time avoid the limitation imposed by linearized perturbative methods. Numerical integration of the Hamiltonian equations to determine the decorrelation among waves caused by the phase fluctuations, and integrating a Boltzmann equation for the surface waves have in the past been limited by limitations in memory and computaticnal capability. *Techniques that can capitalize on the emerging computer technology needs to be developed in such a way as to anticipate the advancements.*

Technical problems associated with describing the evolution of the wind generated field on the sea surface are shared with the phenomenon of turbulence; both are broad band stochastic flow fields. An interesting measure of the statistics of a turbulent flow field can be constructed from a continuous measurement of the velocity field at a point in space. The time trace of the velocity field can be used to determine the distribution of times that the velocity exceeds some preassigned value. This so-called zero-crossing statistic have been used[24] to test for the intermittency of turbulent flows. The theoretical ideas applied to these data analysis assume that the fluctuations in the velocity field have zero correlation time. The effects of a power-law distribution function and/or memory in the fluctuations have not been determined. Just as in the false alarm problem discussed earlier recent results indicate that non-Markovian effects can strongly influence the zero-crossing statistics.[25,26] These ideas have also recently been used to describe the statistical properties of wind-generated water waves.[27]

In Fig. 6 we summarize the above discussion in slightly different language. Firstly, we note that fluctuations in fluid flow cause noise in a sensor return. Secondly, that when these "fluctuations" are caused by ocean waves, which due to nonlinear wave-wave interactions can have non-Gaussian statistics, the noise may require non-standard analysis for its suppression. Thirdly, that even when the ocean waves are relatively benign, the intermittent nature of turbulence may contaminate the sensor return sufficiently to again pose a non-standard signal enhancement question.

STATISTICAL FLUID FLOWS

- FLUID FLUCTUATIONS CAUSE NOISE
  - OCEAN SURFACE WAVES INTERACTING WITH A SURFACE CURRENT
  - LIGHT THROUGH THE TURBULENT ATMOSPHERE
  - SOUND THROUGH THE OCEAN (FILLED WITH INTERNAL WAVES)
  - RADIO WAVES THROUGH THE IONOSPHERE (TURBULENT PLASMA)

- OCEAN WAVES
  - NON-GAUSSIAN NON-EQUILIBRIUM STATISTICS
  - GENERATED BY FLUCTUATING WIND (SURFACE WAVES)
  - NONLINEAR ENERGY TRANSFER

- TURBULENCE
  - INSTABILITIES IN NONLINEAR DYNAMICAL SYSTEMS
  - NONGAUSSIAN STATISTICS
  - INTERMITTENCY
  - LARGE SCALE SPATIAL COHERENCE

Figure 6.

The key issue here is how one handles data streams that contain noise of a fractal nature.[9,10] This consideration in a number of disguised forms, pervades the entire ASW activity. For example the radar return from the sea surface can give intermittent spikes of 10-15 db leading to *intermittency* in the statistics of the radar return. This phenomenon has not as yet been satisfactorily explained.

ASW is a classic example of the signal/noise problem. The surface modulation pattern (signal) is embedded in a complex wind wave field (noise) and the question arises as to the likelihood that a signal of a given strength obtained by scattering a wave (radar from top, acoustic wave from bottom) from the *modulated region* is a *false alarm* rather than the actual detection of a target. Historically it has been necessary to assume that the noise was delta-correlated in time in order to obtain an analytic estimate of the false alarm probability. It is clear that the correlation of the fluctuations of the sea surface (wind generated waves ranging in scales from centimeters to kilometers) that disrupt the modulation pattern can have a persistence time of minutes, thereby invalidating all previous estimates of the frequency of false alarms. A prototype model has recently been developed that yields the exact probability that a signal will achieve a preassigned value in a given time interval for fluctuations having a memory.[25,26]

2.2 Wave Propagation in Random Media

The success of many of DARPA's applied research programs relies on the judicious application of generic models. Such models encapsulate the general features of a wide variety of problems and the investigator fine tunes the model to his particular applications. A number of programs have suffered from the lack of such generic models. Examples that readily come to mind include the disruption of satellite communications, radar imaging and undersea communication.[7] All of these involve wave propagation through media that randomly fluctuates in space and/or time. In dealing with rate processes in such media one has in the past been forced to assume that the spatial and/or temporal fluctuations are delta correlated, i.e., "Markovian," thus restricting the scales of the fluctuations to be much shorter than those of the propagating waves. Although these approximations are often known to be blatantly unphysical, the lack of a generic model with which to guide the analysis has made them necessary, the only alternative being large scale computer calculations, often a requirement even when the approximation is made.[7]

There are several reasons why this work is important scientifically and pertinent to DARPA's interests. The use of wave propagation for scientific purposes, in situations where an intervening random medium is relevant, is widespread. In general, there are two sides to this

coin. On the one hand the fluctuating medium distorts signals and must be understood in order to unfold this distortion as far as possible. On the other hand the presence of this distortion on a signal of known properties provides a method of measuring the properties of the fluctuating medium itself. The removal of distortion dominates in applications involving astronomical observations, both optically through the atmosphere and with radio waves through the ionosphere. The measurement of the intervening medium dominates in acoustic and electromagnetic sounders measuring properties of the atmosphere or the solid earth. Both aspects are present in the use of radio waves in the ionosphere and interplanetary plasma, and sound in the ocean.

Specific applications of interest to DARPA over the long term in this field include applications to HF, VHF, and UHF communications through the ionosphere, between earth stations or from satellites; the accurate pointing of lasers through the atmosphere; the resolution of image systems in satellites; acoustic communications at short range underwater; long-range acoustic detection of submarines; and eventual study of the small-scale inhomogeneities in the solid earth.

At present, the scientific community's ability to predict coherence properties of waves is reasonably good in situations in which the distortion is weak (unsaturated scattering); that is, where the intensity fluctuations are small compared with unit signal strength. The research recently completed at CSND has provided a theoretical basis for predicting the coherence times, lengths, and properties of waves distorted by different media, in situations where the intensity fluctuations are near unity (saturated scattering).[7] One of the important studies that remains is to provide an *independent* test of the properties of the medium deduced from the transmitted waves. One way to do this is by examining the properties of the extreme variations in the amplitude, phase and intensity of the transmitted wave.[12]

Masoliver et. al.[25,26] have recently made a breakthrough in the area of extreme statistics of stochastic processes that enables one to analytically determine the effects of non-Markovian fluctuations on a prototype problem. Their approach is one involving stochastic path integrals. The problem they have solved is that of a one-degree-of-freedom rate equation yielding a trajectory in a non-Markovian one-dimensional medium. The method so far seems to be quite general and may be extended in a number of different directions. The most obvious extension is to many-degree-of-freedom systems, and they foresee no conceptual difficulties with these generalizations.

They have applied their formalism to three specific problems: diffusion in a continuous medium, diffusion in a discrete medium, and Brownian motion.[25,26] The effects of correlated

WAVE PROPAGATION THROUGH CORRELATED RANDOM MEDIA

WAVE PROPAGATION

- SOUND THROUGH THE OCEAN
- LIGHT THROUGH THE ATMOSPHERE
- RADIO WAVES THROUGH PLASMAS

PREVIOUS THEORIES

- MOMENT EVOLUTION EQUATIONS
- PATH INTEGRALS
- RELIANCE OR CENTRAL MOMENT PROPERTIES (NOT UNIQUE)
- DELTA CORRELATED TIMES AND DISTANCES

NEW TECHNIQUES

- PATH INTEGRAL (GENERIC MODEL)
- NOT RESTRICTED TO DELTA-CORRELATED FLUCTUATIONS IN MEDIA
- EXTREMAL PROPERTIES OF FLUCTUATIONS ARE MORE SENSITIVE PROBES OF MEDIA STRUCTURE THAN ARE CENTRAL MOMENTS

Figure 7.

fluctuations are determined to be quite dramatic when they contrast these results with those obtained for delta-correlated fluctuations. For example, the capture rate of a *diffusing particle* at a trap may be changed by an order of magnitude by a non-Markovian environment. An even more dramatic difference in capture rates may occur when the diffusing particle is under the influence of a binding potential. This theoretical approach takes advantage of the extreme variability of the exact solution to the dynamic equation. The extension of the technique to two dimension will allow the study of the transmitted wave in the saturated region. Instead of using the traditional statistical measures, those being the central moment and correlation function analysis, one can directly investigate the wings of the distribution of the fluctuations. The tails or wings of the distribution are much more sensitive probes of the structure of the medium than are the central moments.[12]

3. Organization of Complex Programs

3.1 Complex Programs and Dimensionless Constants

One of the most fruitful strategies for modeling complex dynamical systems has been the use of dimensionless constants. Long before there were computers to numerically integrate complicated equations like those due to Navier and Stokes for describing fluid flow, scientists used dimensionless constants for the design of experiments involving small scale physical models of full sized objects. The design of ships, dams, harbors, canals, airplanes, etc. would have been impossible with out such scaling experiments. In fluid problems one is often concerned with the Reynolds number, which indicates the ratio of the dominant spatial scale in the fluid flow to the viscous dissipation scale length. The Reynolds number is of central importance in determining the transition from laminar to turbulent flow. In ship design, on the other hand, it is the Froude number that is of central importance since the drag on a ship, to first approximation scales with the Froude number.

This mention of dimensionless constants is important because in large part research programs have developed around projects that rely on a single such constant. We have mentioned that ship design relies on the Froude number, whereas turbulence is centered on the Reynolds number. Thus programs centered on ship design (drag)) turbulence are "simple" in that the scope of the activity is relatively well defined. A complex, but decomposable, program is one that can be segmented into a number of weakly interacting elements each of which is governed by a different dimensionless constant. Consider the ASW program; the transport of energy in the

wind-generated field of sea waves is characterized by a Kubo number, the properties of the wind field by a Reynold number those of the internal wave field are determined by a Richardson number, etc., etc. Any of these constants may be large for the particular phenomenon it controls, say, the Reynolds number for the air flow may be large so that the wind is turbulent, but its value does *not* determine the coupling with the adjacent fluid, i.e., the sea surface. The air-sea coupling parameter is affected by the fluctuations in the wind, but it does not directly depend in the wind Reynolds number.

By contrast a complex, but undecomposable, research program is one in which the dimensionless constant of one process directly influences that of another process. Stated differently, the observed phenomenon is not dominated by a single dimensionless constant, but rather depends equally on two or more of them. The problem of wave propagation in a random medium is of this form since the medium fluctuations enter the index of refraction and therefore the dependence of the wave on the Reynolds number of the medium is amplified. Similarly, the disruption of the wave depends on the correlations in the medium fluctuations. Thus the Reynolds number and a suitably scaled correlation time and correlation distance define three dimensionless parameters that determine the statistical properties of the medium and therefore those of the propagating wave. These dimensionless parameters in turn are the macroscopic representation of the microdynamics of the underlying medium. No research program that ignores all but one of these parameters can hope to succeed in properly describing this phenomenon. For example the theory of Tatarski[6] has only had limited success in describing wave propagation in media where the fluctuations give rise to strong scattering, or in which the scattering is correlated over long distances or times. The more recent state-of-the-art work done at CSND although superior to the earlier work of Tatarski still relies on the Markov approximation for the medium fluctuations.[7]

As more dimensionless constants become important in a research program the opportunity for the individual scientist to make a significant contribution is reduced. Consider for example the problem of weather forecasting, involving as it does atmospheric dynamics, heat exchange at the ocean surface, diurnal cycles, etc.[12] No one of these elements could be excluded from a comprehensive model of weather predictability. This situation motivated the National Science Foundation to form a division called the Global Atmospheric Research Program (GARP) whose mission has been to support those areas of research that directly influence the problem of the predictability of the weather. This is a complex program in the category 2 sense, in which the program managers support a wide variety of research activities but each activity has a specific long term research objective contributing to the understanding of weather predictability.

Research programs at DARPA have a mission orientation that is generally lacking at NSF. Since the feasibility of achieving a mission goal can usually be identified more quickly than can the realization of a scientific one, DARPA's research programs tend to be viable for the order of three to five years, after which time those programs that are successful are picked up by one of the services and those that are not become milestones for other programs. Thus the question arises as to how to design a complex program that has the required level of integration for a DARPA activity without changing the time scale of the research. The recent University Research Initiatives (URI) of DARPA stressed this integrative component, blending scientific excellence with potential technological impact. The organizational design of the research activity in a given topic area of the URI was left in the hands of the responding institutions. It would better serve the interest of DARPA and save time on the part of the proposing scientists if there were a master organizational scheme that provided maximum flexibility to both DARPA management and the research scientists, while at the same time insuring the scientific quality of the research. One such plan is outlined below.

3.2 Organizational Plan for Large Scale Complex Technological Programs

The program liaison with DARPA's program manager will be the Principle Investigator called the Project Director. The Project Director will be responsible for the coordination of the scientific and financial directions of the program and will bear the final decision making responsibility for all aspects of the program.

The objective of the program will be realized through the formation of three committees to be coordinated by the Principle Investigator [cf. Figure 8]:

1) The *External Scientific Advisory Panel (ESAP)*, whose members will be chosen from the academic, industrial and government scientific communities to provide the best available advice and guidance from outside experts in the fields of programmatic interest.

2) The *Technological Transfer Panel (TTP)*, whose members will be chosen from government and industrial settings to provide the best available advice and guidance on the technological implications of the scientific research particularly for the services. It is suggested that the chair of this committee be the DARPA contract monitor for the program.

3) The *Overseer Committee* consists of the senior investigators in the program and will oversee the allocation of funds and personnel, the purchase of equipment, schedules and the logistics of all activities of the program in addition to the overall research direction of the

COMPLEX PROGRAM MANAGEMENT: STRUCTURE, COMMITTEES RELATIONSHIPS AND SCHEDULES

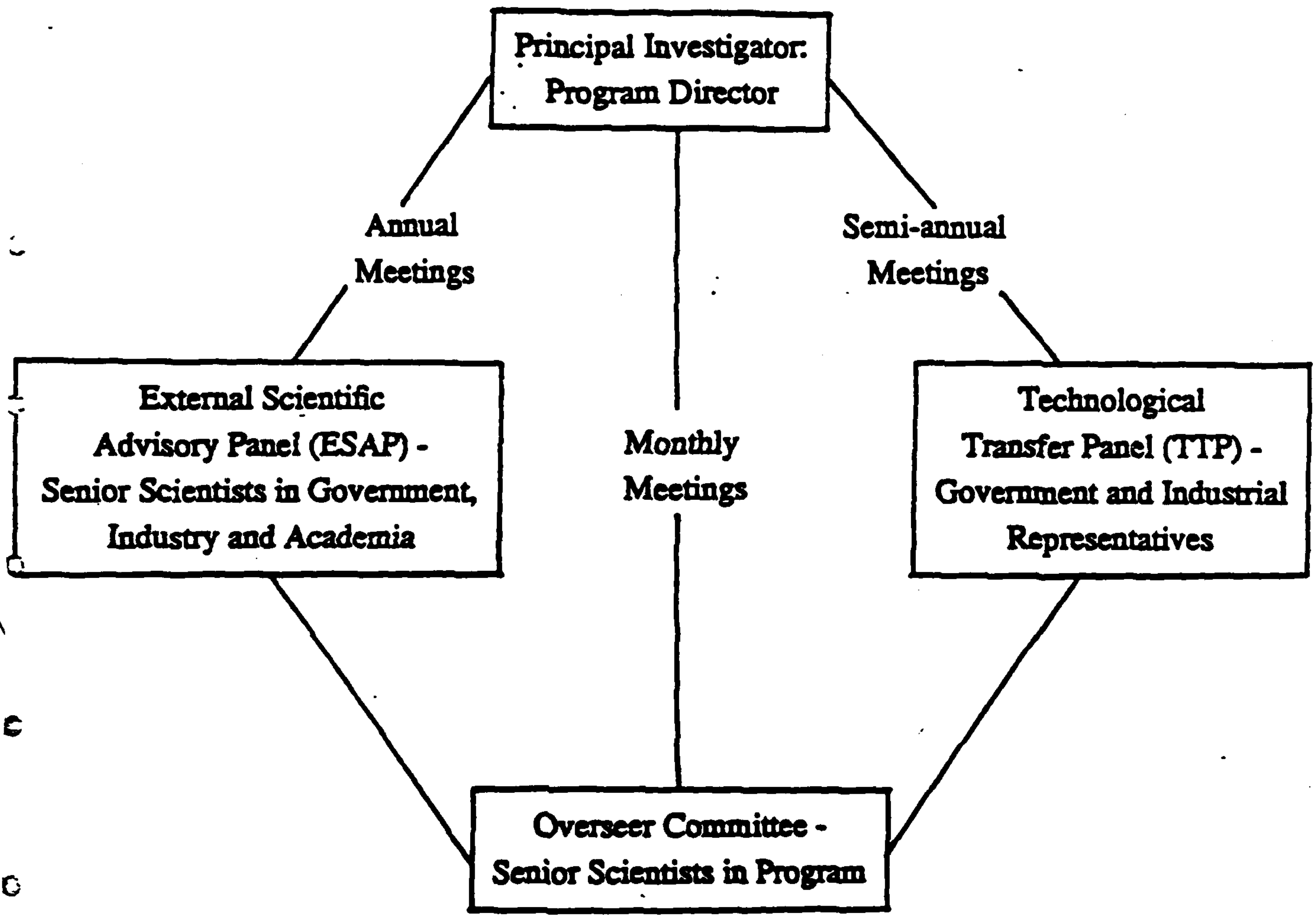


Figure 8.

program.

The ESAP committee will meet annually, the TTP committee will meet semiannually, and the Overseer Committee will meet at least once a month (see Figure 8).

The members of the External Scientific Advisory Panel (ESAP) should reflect the mixture of academic, industrial and defense perspectives on the programmatic problems of interest to DARPA (cf. Fig. 9). The function of the ESAP is to provide scientific guidance and perspective from outside the program, to critique both the scientific quality of the research and its compatibility with the stated directions of the program and to review the semiannual report provided by the Program Director. The Panel will meet annually and will prepare a report on their findings. This report will be an important component in the decision-making processes that determine the ongoing direction of the program. The Program Director will provide the members of ESAP with copies of all printed material associated with the program including manuscripts, technical reports, and copies of the semi-annual reports. All these materials will be provided in a timely fashion to enable ESAP members to respond at the yearly meetings.

The purpose of the Technological Transfer Panel (TTP) is to provide the program with a broad perspective of the possible technological impact of the research (cf. Fig. 10). This panel will be constituted in such a way that it will be able to advise the program on the mutual implications of the research performed within the program and elsewhere, and on the long-term consequences of this research particularly for the armed services. Although scientific breakthroughs are unpredictable, technological advancement is not, and TTP can help to stimulate the kind of scientific activity necessary for these technological advances. It can also, in this context, facilitate the exchange of short and long term visitors between the program and government/industrial laboratories. We suggest that the panel be chaired by the DARPA Contract Monitor for this program.

The TTP will meet semiannually. These rather frequent meetings are intended to offset the presumed relative unfamiliarity of the program scientists with broad-based technological issues. Under the guidance of the TTP committee the relevant technological questions can be addressed in the scientific context of the program. The TTP committee will prepare an annual report which will be distributed to the Overseer Committee, to the ESAP committee, to DARPA and to the other interested parties.

The duties of the members of the Overseer Committee include:

FUNCTIONS OF EXTERNAL SCIENTIFIC ADVISORY PANEL (ESAP)

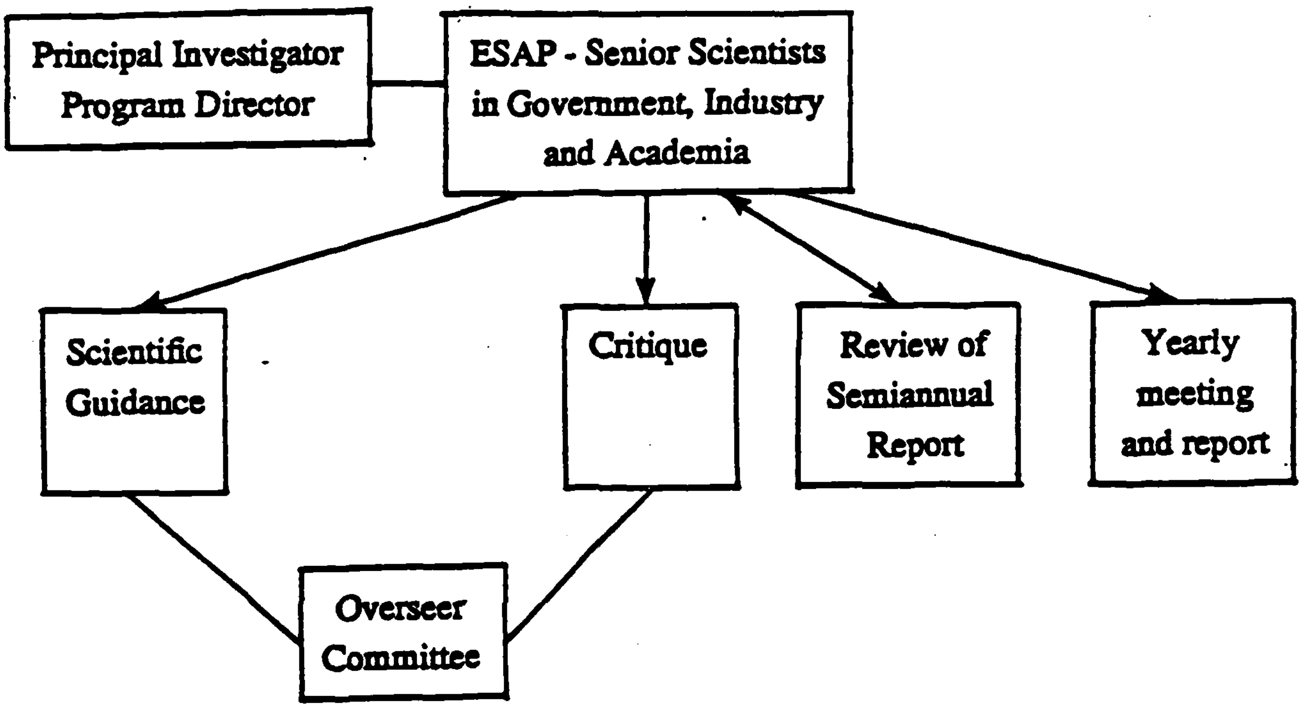


Figure 9.

FUNCTIONS OF TECHNICAL
TRANSFER PANEL (TTP)

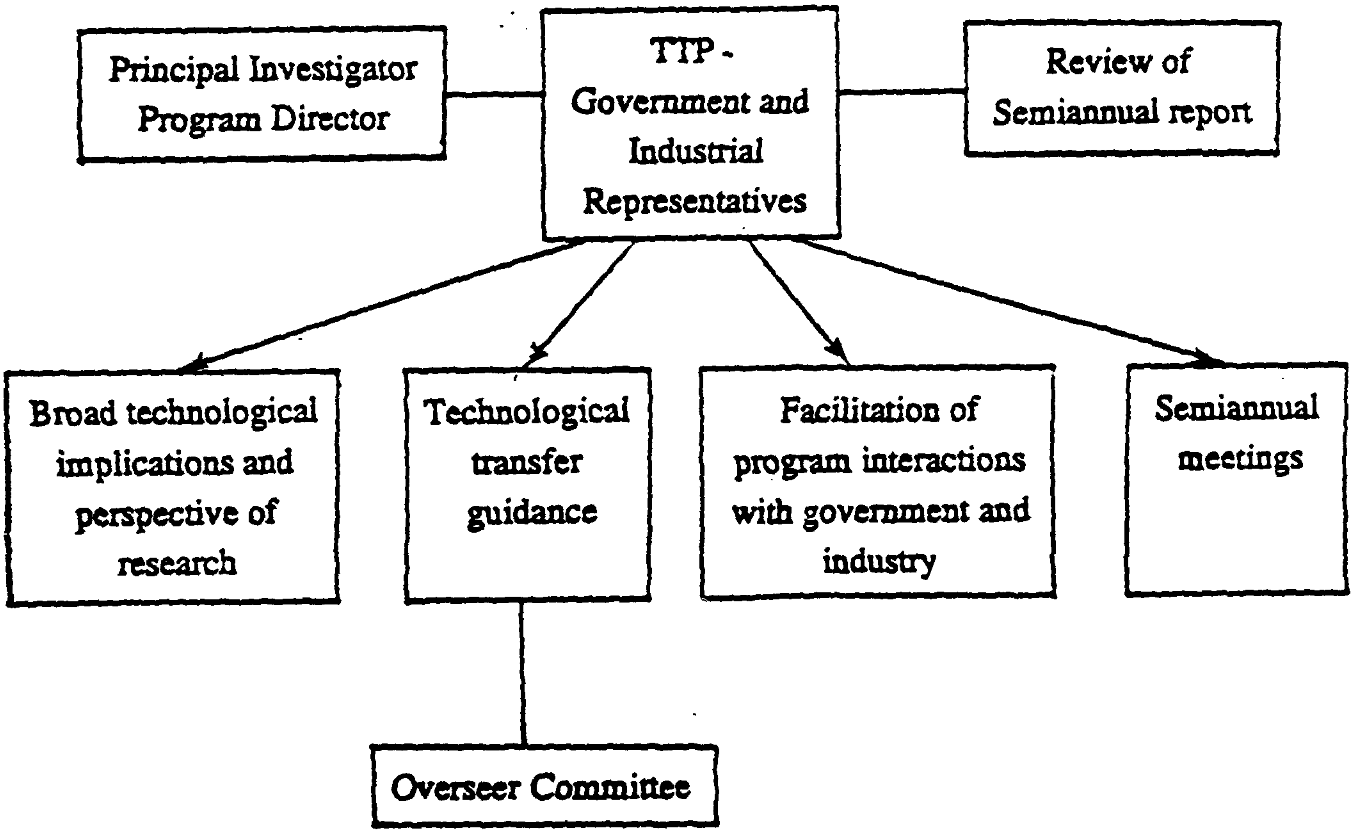


Figure 10.

i) To keep abreast of the scientific progress in the individual projects of the program and to assess the compatibility of this progress with the stated objectives of the program. The Committee must of course be cognizant of recommendations made by the ESAP and the TTP.

ii) To make recommendations for the allocation of common recourses including funds for equipment, personnel, travel, seminars and workshops, and supplies.

iii) To oversee the arrangement of workshops and seminars.

iv) To oversee the initiation and development of interactions with industrial and government scientists with those in the program.

v) To interact with the ESAP and TTP committees, inform them of progress and problems, and communicate their advise to the program at large.

vi) To provide the Program Director with individual semiannual summaries.

The internal allocation of resources is a very important and often sensitive program function. Although broad categories would be specifically assigned in a proposal, their implementation must be safeguarded and requires priorities to be set within the categories. Therefore, taking into account the continual input of the Overseer Committee, the Program Director would have the final decision responsibility for all the allocation of resources.

The program will foster both short and long term visits to and from government/industrial research institutions. These exchanges will facilitate the dissemination of research results and will stimulate an appreciation for broader-based research goals and technological applications on the part of participation in the program.

Several types of workshops might facilitate internal communication and the smooth running of the program:

i) Internal review workshops involving mostly program participants These workshops will be devoted to a general review and critique of the work of the program over the previous year. These sessions will be coordinated with the meetings of the ESAP and TTP committees.

ii) Internal "school" workshops for junior scientists, of a pedagogical nature with the senior members of the program preparing three or so hours of lectures on their work.

iii) Open thematic workshops devoted to a general forum on particular programmatic themes. The speakers will primarily be from the outside of the program and will include a good representation of scientists from industrial and government laboratories. These will be open brainstorming sessions to enlarge the scientific horizon of the program for future work. These

workshops will be organized in close collaboration with ESAP.

iv) Technological forecasting workshops devoted to the enlargement of the program sights of the potential applications of the research conducted under the program. These workshops are to be developed and run in close collaboration with TTP.

The present managerial scheme only effects the highest organizational level of the program. For truly complex systems the full impact of this new approach will only be felt if the structure is replicated further along the hierarchy. It would seem advisable: 1) to make each senior scientist in the Overseer Committee a Director of their individual scientific project as shown in Figure 11 and 2), to form an Overseer Committee in each project area from the senior scientists within that group and to form both an ESAP and a TTP for the project. In this way not only does each project derive the same benefit from the SCOMPLEX as does the overall program, but that benefit is amplified due to the self-similar organizational structure.

In Figure 12 we indicate how these separate elements at each level of the hierarchy are to be used as a self-assessment feedback loop in which information regarding all activities of the program are critiqued. The senior investigators in the program will provide the program director with semiannual summaries of the research being done under their direction. (These are the same scientists that collectively form the Overseer Committee). These summaries will be synthesized with the recommendations of the ESAP and the TTP by the program director into semiannual reports. These reports will be supplied to the members of ESAP, TTP and the Overseer Committee for comments. In this way all aspects of the program are monitored by DARPA, the outside advisors and most importantly by the scientists within the program. This format for program managing inhibits the foundation of separate fiefdoms by individual investigators and promotes collaborative activities.

The present organizational plan has been devised with category 2 research in mind. Consider the problem of wave propagation in random media. The aspects of the program could well consist of three or four of the applications mentioned that are of interest to DARPA in conjunction with a modeling activity that would focus on applying new mathematical techniques to the problem. The program would then have four or five interrelated research activities that would be coordinated by the Program Director and the Overseer Committee. The organizational plan would promote joint activities that would highlight the common features of laser propagation in the atmosphere, radar propagation in the ionosphere, sound propagation in the ocean, etc. The development of the in-depth rather than superficial similarities among those phenomena would

COMPLEX PROGRAM MANAGEMENT STRUCTURE

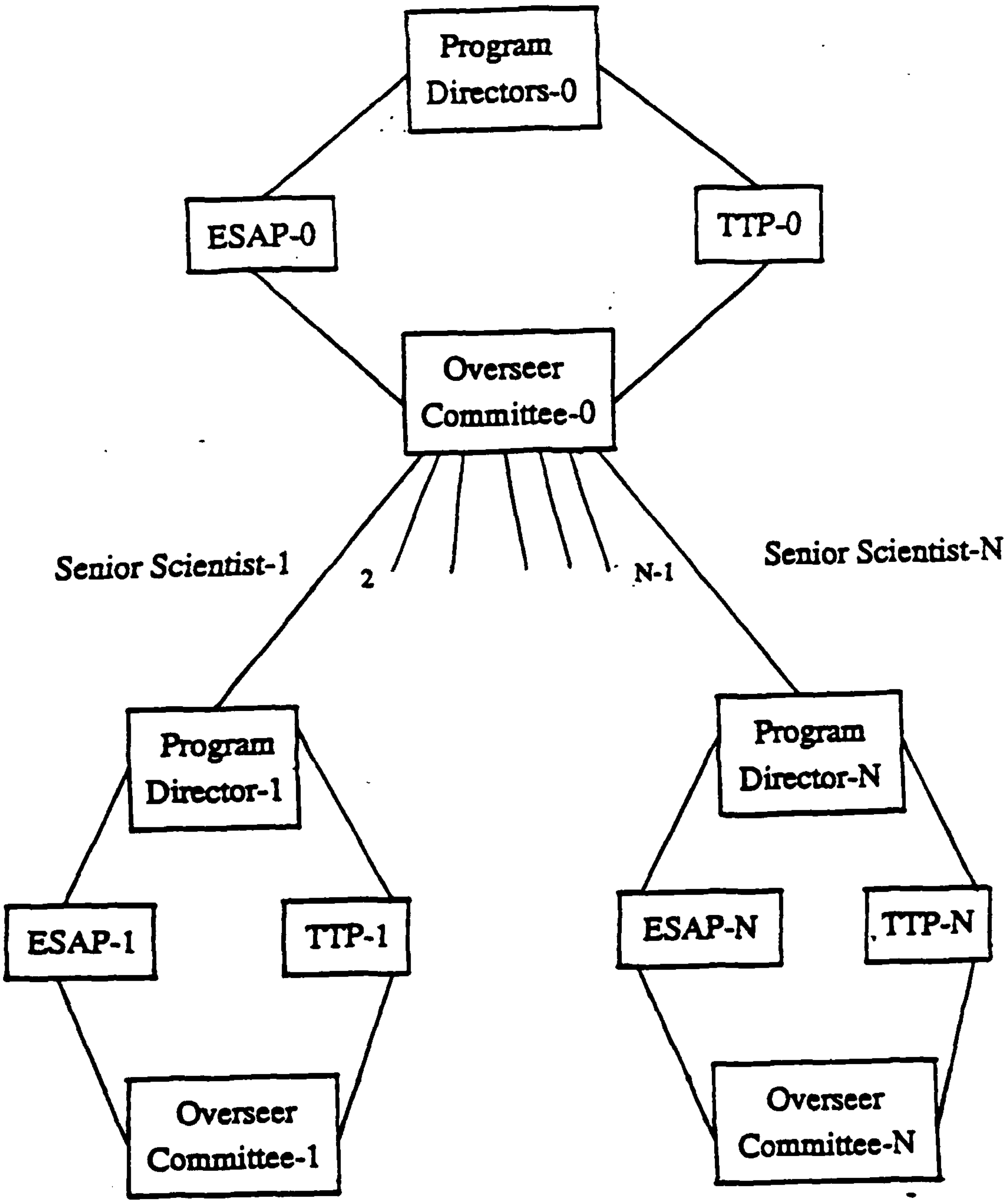


Figure 11.

SELF-ASSESSMENT PROCEDURE

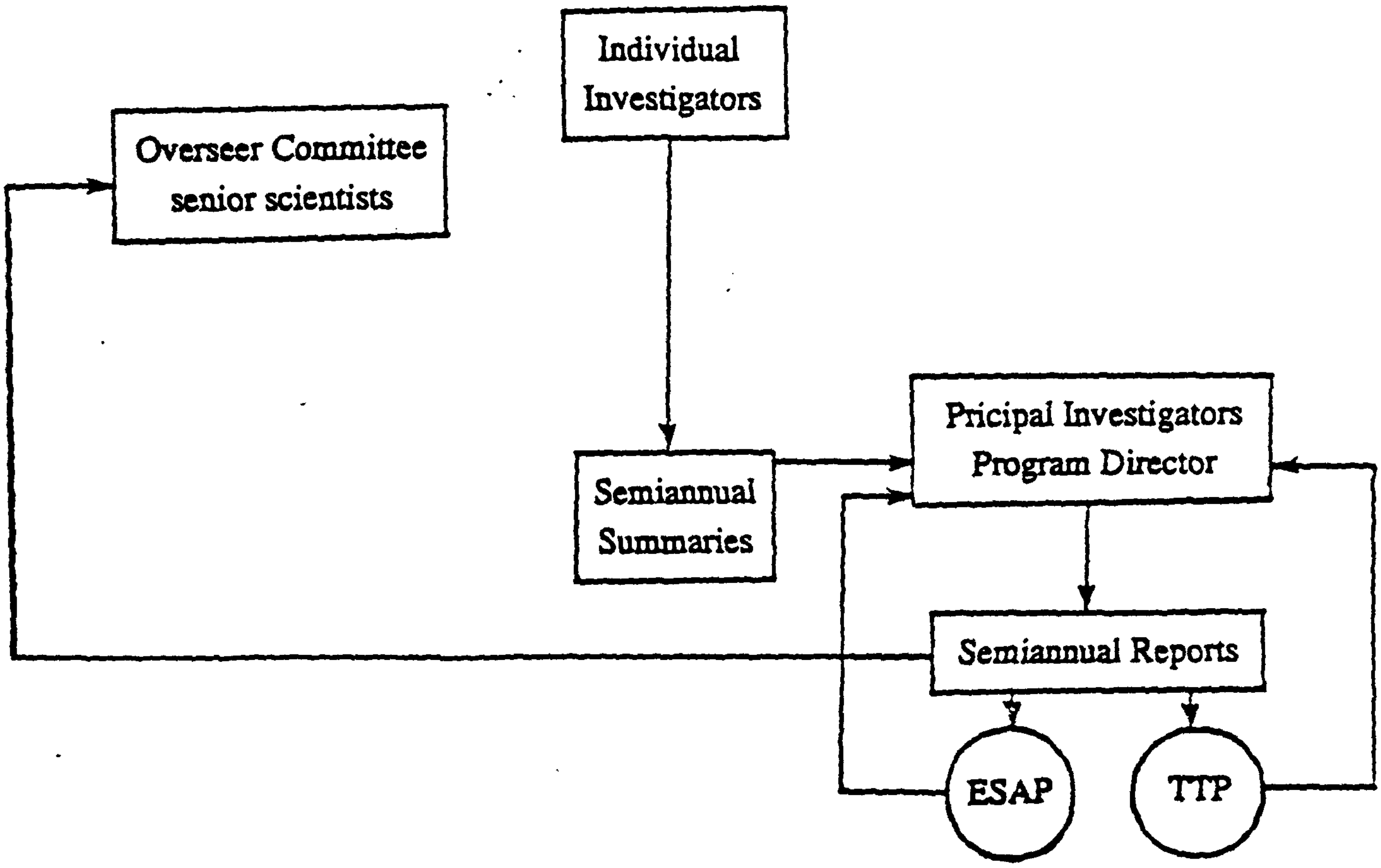


Figure 12.

form the basis for a generic mathematical model of wave propagation that includes correlated fluctuations in the medium. The phenomenology of fluctuations in the atmosphere, the ionosphere (plasma) and the deep ocean as well as their dynamics description must be developed in a coordinated way in order to understand how they influence the propagation of laser, radar and acoustic waves, respectively.

Category 1 research can also benefit from the present organizational plan since it would foster stronger research ties among the various research activities. For example, in ASW the modeling of radar scattering from the sea surface would be greatly improved by the collaboration of physical oceanographers, scattering physicists and data processors. The blending of the state-of-the-art development in each of these areas would be insured by the *self-assessment procedure*. The mutual benefit derived from these formally structured interactions would nurture informal collaboration and accelerate the resolution of *difficult* technical problems.

4. SUMMARY

Herein we have proposed the formation of a strategy for complex organization modeling, planning and experiment (SCOMPLEX). The strategy is an organizational plan for the monitoring and control of large scale complex technological programs. We have divided complex programs into those that can be segmented into isolated (weakly interacting) units (category 1) and those that cannot (category 2). ASW has been used as an example of a category 1 research program [cf. Figure 2] and wave propagation in random media has been used as an example of a category 2 research program [cf. Figure 3]. The decision as to the appropriate category into which a proposed project fits, can be made based on an examination of the number of dimensionless constants necessary to describe the basic problem and whether or not these constants are independent. Once it has been decided if the project is category 1 or 2, the procedures outlined under SCOMPLEX may be adopted or not. The key to the organizational plan is the self-assessment procedure indicated in Figures 8 and 12, since it is through this procedure that one decides if the program is satisfactorily accomplishing its stated research goals. If it is not realizing these goals, then the mechanism is in place for implementing modifications that will redirect the research effort. Further, the self-similar character of the organizational scheme suggested in Fig. 11 insures a consistent level of success at each level of the hierarchy.

REFERENCES

1. PREDICTABILITY OF FLUID MOTIONS, Editors G. Holloway and B.J. West, AIP Conf. Proc. 106 (1984).

2. A.S. Monin, WEATHER FORECASTING AS A PROBLEM IN PHYSICS, MIT Press, Cambridge, Massachusetts (1972).

3. MATHEMATICAL METHODS IN HYDRODYNAMICS AND INTEGRABILITY IN DYNAMICAL SYSTEMS, Editors M. Tabor and Y.M. Treve, AIP Conference Proc. 88, (1982).

4. B.J. West, AN ESSAY ON THE IMPORTANCE OF BEING NONLINEAR, Lecture notes in Biomathematics 62, series editor S. Levine, Springer-Verlag, Berlin (1985).

5. B.J. West, DEEP WATER GRAVITY WAVES, Lecture Notes in Physics 146, Springer-Verlag, Berlin (1981).

6. V.I. Tatarski, THE EFFECT OF THE TURBULENT ATMOSPHERE ON WAVE PROPAGATION, (Keter, Jerusalem, 1971).

7. "Theoretical Studies and Data Analysis of Wave Propagation in Random Media" final report (1985) for DARPA contract MDA 903-83-C-0515.

8. A.S. Monin and A.M. Yaglom, STATISTICAL FLUID MECHANICS VOLS. I AND II, MIT Press, Cambridge, Mass (1975).

9. B.B. Mandelbrot, 1982, THE FRACTAL GEOMETRY OF NATURE, W.H. Freeman, San Francisco.

10. M.F. Shlesinger, B.J. West and J Klafter, 1986, "A Lévy Walk Representation of Turbulent Diffusion" submitted to Phys. Rev. Lett.

11. H.G.E. Hentschel and I. Procaccia, Physics Review A 28, 417 (1983)

12. K. Lindenberg and B.J. West, "The Biggest, the Best and Other Such Considerations," J. Stat. Phys. 42, 201 (1986).

13. S.M. Flatté, Proc. IEEE 71, 1267 (1983).

14. K.G. Wilson, Scientific American, 241, 158 (1979).

15. Proc. of a SYMPOSIUM ON FRACTALS IN THE PHYSICAL SCIENCES, editors M.F. Shlesinger, B.B. Mandelbrot and R.J. Rubin, J. Stat. Phys. 36 (1984).

16. B.D. Hughes, E.W. Montroll and M.F. Shlesinger, J. Math Phys. 28, 111 (1982).

17. B.D. Hughes and M.F. Shlesinger, J. Math Phys. 23, 1688 (1982).

18. M.F. Shlesinger, J.Klafter and Y.M. Wong, J. Stat. Phys. 27, 499 (1982).

19. B.D. Hughes, E.W. Montroll and M.F. Shlesinger, J. Stat. Phys. 30, 273 (1983).

20. E.W. Montroll and M.F. Shlesinger, Proc. Nat'l. Acad. Sci. USA 79, 3380 (1982).

21. B.J. West and V. Seshadri, Physica 113A,203 (1982).

22. B.J. West and V. Seshadri, Proc. Nat'l. Acad. Sci. USA 29, 4501 (1982).

23. See e.g. K.M. Watson, B.J. West and B.I. Cohen, J.of Fluid Mech. 27, 185 (1976).

24. K.R. Sreenivason, A. Prabhu, and R. Narasimka, J. of Fluid Mech. 137, 51 (1983).

25. J. Masoliver, K. Lindenberg and B.J. West, "First-Passage Times for Non-Markovian Processes," Phys. Rev. A33, 2177 (1986).

26. J. Masoliver, K. Lindenberg and B.J. West, "First Time for Non-Markovian Processes: Correlated Impacts on a Free Process", Phys. Rev. A34, (1986).

27. B. J. West and K. Lindenberg, "State-Dependent Fluctuations in Open Systems: Simple Models" to appear in STUDIES IN STATISTICAL MECHANICS XIII, ed., J. Lebowitz, North-Holland, Amsterdam (1986).